

Going public

Should scientists let the public help them decide how government research funds are spent? Yes they should, because the consequences are to be welcomed, not feared.

Science communication, *circa* 1600: discussions with the public, according to one prominent researcher, are little better than listening to the “maunderings of a babbling hag”. So said William Gilbert, a pioneer of research into electricity and magnetism.

Today’s scientists are, at least in the main, a more open-minded bunch. But the prejudices and fears that underlie Gilbert’s remark have not entirely gone away, as reactions to some new initiatives show.

Take last month’s report by Demos, a UK political think-tank. For many researchers, it will make frightening reading. The left-leaning Demos makes the first coherent call for ‘upstream engagement’ — the involvement of non-specialists in setting research priorities. British scientists have seen the public swayed by misleading media coverage of genetically modified (GM) food and vaccines. For them, the proposal must seem close to giving the lunatics the keys to the asylum.

Such concerns will not be restricted to Britain: environmental organizations across Europe are committed in practice to ending research into GM crops. Some religious groups in the United States would end research involving human embryos if they had the power to do so. And it would be impossible to develop safer and more efficient nuclear power stations, which will probably be needed to tackle climate change, if anti-nuclear groups have too much influence on research policy.

Yet there are good reasons why scientists should ignore these fears and embrace upstream engagement. On an ethical and political level, the research community has no right to reject public involvement outright. Taxpayers fund research, buying themselves the right to help shape its course. Objecting to public involvement would simply undermine the current enthusiasm shown for science funding by some governments, such as those in the United States and Britain.

Balance of power

There is also plenty of evidence to suggest that upstream engagement, if managed properly, will not bring an end to any area of research. Such engagement is already being quietly and usefully practised in the research-charity sector, where the trustees of many funding organizations are non-scientists. And the slew of new initiatives being proposed for the public sector involve giving the public less power than the trustees, and certainly not a veto over research spending.

When worrying about engagement, British researchers may also be swayed too much by the GM fiasco, in which propaganda put out by environmental groups and the biotechnology industry made public debate extremely difficult. But other exercises have proved less combative and more fruitful. The Natural Environment Research Council, for example, last year ran public consultations on a new research programme. It led to a new theme — the sustainable management of marine bioresources — being added to the programme.

Get the process right, and other consultations could produce equally meaningful input. No one wants to haul people off the street and make decisions based solely on questionnaires. There are numerous mechanisms for engaging the public, from citizens’ juries to consensus conferences and deliberative mapping processes. The details vary, but all involve giving non-specialists access to a range

of different perspectives on a particular topic, and allowing them to develop their own recommendations through structured discussion. Sociologists say that the techniques need to be evaluated to see which works best, but that’s no reason not to start now.

Funding bodies are the obvious target for engagement exercises. In the United States, the National Institutes of Health (NIH) faces increasing lobbying from advocacy groups, often representing the needs of patients with a specific disease, who want the agency to do less basic research and more drug development. Public engagement could help the NIH bolster its efforts to incorporate a broader range of views into its decision-making processes.

Nothing to fear

In many European nations, there is little call for upstream engagement. But Britain, where a lack of public trust in science is perceived as a serious problem, is a notable exception. Not all of the country’s funding bodies have taken this on board. The Biotechnology and Biological Sciences Research Council, which is setting up a permanent committee of non-scientists to advise on strategy, leads the way. But the Engineering and Physical Sciences Research Council (EPSRC) has lagged behind. This is worrying, as the council funds research in nanotechnology, an area of science that could one day transform everything from drug delivery to computing.

Judging by the few consultations that have already been run on nanotechnology, the EPSRC should not fear public involvement. Non-specialists tend to reject the call for a moratorium on nanotech research made by one more extreme environmental group. Instead, they suggest that environmentally useful applications, such as new solar-power systems, should be made more of a priority. More work on the environmental impact of nanoparticles is another common request — a call echoed by July’s report on the same topic by the Royal Society and the Royal Academy of Engineering.

Upstream engagement is no panacea. On its own, it won’t solve Britain’s crisis over trust in science. Nor will it resolve thorny questions about what types of science are worth pursuing, and which should be avoided because of links to technology such as weapons of mass destruction. But it is worth doing — provided that all involved consider two points before beginning.

First, the processes must be long-term and properly funded. Money spent on engagement is often diverted from basic research. So if governments are serious about upstream projects, they should talk to research agencies about how to ring-fence money to run the consultations. In Britain, this is likely to amount to a few million a year across all sciences, a fraction of a per cent of the total science budget.

More importantly, funding organizations must make a genuine commitment to react to the results of engagement processes. This doesn’t mean simply accepting the outcomes; research councils should clearly remain in ultimate charge of priority setting. But for the process to be meaningful, funders must explain why they choose to accept some pieces of advice and reject others. The UK government ran a public debate on genetic modification last year and is widely believed to have ignored the results — something only a little less offensive than talking about babbling hags. ■





Russian roulette
Academy labs
anxiously await
reform proposal
p887



Star power
Schwarzenegger
backs California
stem-cell initiative
p888



Cross words
Author and editor
square up over book
on rebellion
p889



Ravenous horde
Plague of locusts
threatens North
African agriculture
p890

Biologists seek consensus on guidelines for stem-cell research

Erika Check, Washington

How would you know if a human brain was trapped in a mouse's body?

That might sound like a question for science-fiction writers, not scientists. But prominent biologists discussed it at a meeting at the US National Academies, held in Washington DC on 12–13 October, in an attempt to frame some ethical guidelines for research on human embryonic stem cells.

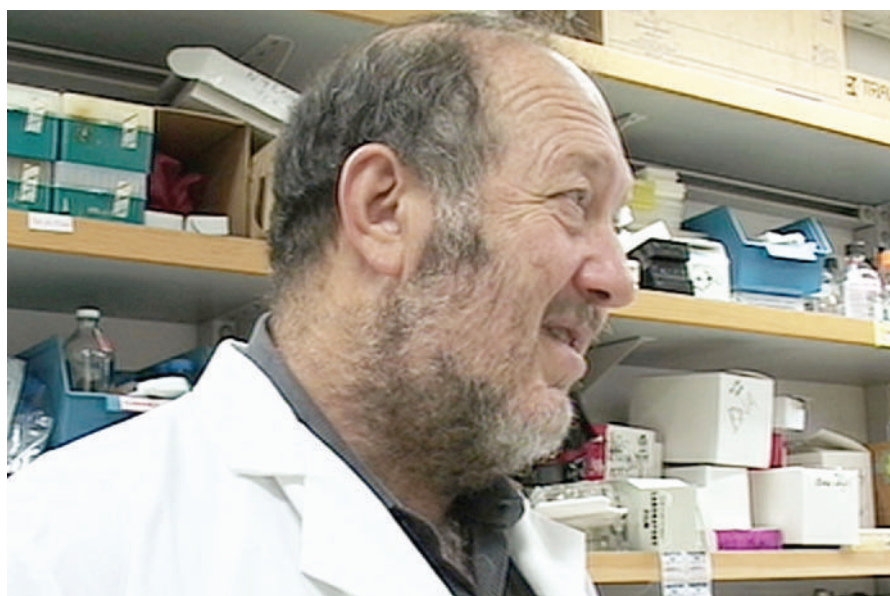
The academies plan to issue guidelines next February. And in the absence of any US government regulation of the expanding field — much of which is likely to be funded privately, owing to restrictions on its public funding — the suggestions are set to play a critical role in determining what experiments are done with stem cells.

Researchers at the meeting agreed on a lot: that the use of human embryonic stem cells to produce a baby should be banned, for example, and that stem-cell researchers should adopt guidelines to reassure the public that their work is ethically sound. But they differed on how to handle chimaeras, which mix cells and DNA from different species.

Many researchers think they could gain insights into human development and disease by transferring human embryonic stem cells into embryos, fetuses or newborn animals of other species. But the research poses uncomfortable questions about the boundary between man and mouse — or human and monkey, dog or sheep.

Irving Weissman, a cancer biologist who leads the privately funded Institute for Cancer/Stem Cell Biology and Medicine at Stanford University in Palo Alto, California, said that human-to-mouse stem-cell transplants have already proved invaluable to medicine. By seeding developing mice with human embryonic stem cells carrying genetic predispositions to disease, Weissman said, scientists could study how such diseases develop in animals. "Every level of research I do on stem and progenitor cells is enriched by making human-mouse chimaeras," he said, adding that much of it was "only possible" using the chimaeras.

Scientists could even construct a mouse whose entire brain was made of human-



Mixed blessing: Irving Weissman supports experiments using chimaeras of human and mouse cells.

derived cells, Weissman said, although he has not performed such an experiment.

But others were concerned at the idea of a mouse with human brain cells. Brigid Hogan, a developmental biologist at Duke University in Durham, North Carolina, said that such research is troubling from an animal-welfare perspective, and becomes more so as the mouse brain becomes more human-like. She proposed that researchers adopt a scheme in which experiments would be allowed, disallowed, or merit "deeper scrutiny".

Order sought

"What I would like to see is some sort of hierarchy of regulation for experiments in which a large proportion of neuronal centres of the brain were going to be derived from human cells," Hogan said. "I would like to see evidence that those experiments could not be done in any other way."

Weissman disagreed, saying that he had not heard a good argument for any ban. "I'm not saying we should be unrestricted," he said. "But I haven't heard the reason yet for restricting certain experiments." Weissman said, for instance, that the 'yuck factor' is no reason to ban an experiment — but that

health dangers to mother and child could justify a ban on reproductive cloning.

Other researchers proposed that scientists set up a body similar to the Recombinant DNA Advisory Committee, which is run by the US National Institutes of Health (NIH), to review questions on stem-cell research. But it is not clear where such a body would be located. The NIH washed its hands of much stem-cell research after President Bush decided not to publicly fund work on cells derived after 9 August 2001. And the research is too far from the clinic for the US Food and Drug Administration to get involved. That leaves a hole for scientists, who are not sure what the law permits them to do, and lack guidance on their work's impact on public opinion.

The stem-cell committee is co-chaired by cell biologist Richard Hynes of the Massachusetts Institute of Technology and bioethicist Jonathan Moreno of the University of Virginia. Its recommendations will not be law, but Len Zon, president of the International Society for Stem Cell Research, said his group will help spread them through the scientific community. "If posed in the right way, they would actually speed the research," Zon said. ■

Infertility specialists counsel caution over frozen eggs

Helen Pearson, Philadelphia

Research advances that enable women's eggs and ovarian tissue to be frozen for later use in assisted reproduction are not yet suitable for widespread adoption, says a report from the American Society for Reproductive Medicine (ASRM).

The methods "are by no means ready for prime time," Mark Fritz of the University of North Carolina at Chapel Hill, head of an ASRM committee that advises members on clinical practice, told the society's annual meeting in Philadelphia this week. Fritz said that doctors need to collect far more data on the health of children born through these methods before they become widely used.

A woman's fertility generally declines from puberty onwards, and more steeply from her thirties, as eggs are ovulated, die or deteriorate. Women with cancer and other illnesses can be rendered infertile much earlier if their eggs are damaged by chemotherapy or radiotherapy.

One way round this is to remove and freeze eggs before the treatment, for use later for *in vitro* fertilization. But doctors are still working out how to reliably prevent formation of damaging ice crystals in the eggs, and there have been fewer than 100 live births through this method. At least two US companies, however, offer women the chance to freeze their oocytes.

Another option still at the experimental stage is to remove and freeze entire slices of ovarian tissue, containing many eggs, and reimplant them later. Last month, doctors in Belgium announced the birth to a former cancer patient of the first baby produced by this technique.

The ASRM report 'Ovarian tissue and oocyte cryopreservation', is also published online this month in *Fertility and Sterility* (82, 993–998; 2004). It agrees that the techniques should be offered under strict supervision to cancer patients or other women who will be rendered infertile by medical treatment and who have few other options if they want children.

Before such techniques are offered to infertile but healthy women, the committee recommends that more data be gathered on any adverse effects on a woman's health or the risk of abnormalities or long-term health problems in her children. "Worldwide experience is simply insufficient at this time," Fritz says. The ASRM guidelines are not binding, but those in the field say that most US fertility clinics are expected to abide by them. ■

Early success claimed for Zerhouni's NIH roadmap

Meredith Wadman, Washington

It is just over a year since Elias Zerhouni, director of the National Institutes of Health (NIH), announced a far-reaching plan to transform key areas of biomedical research in the United States. And last week, the NIH director declared that his "roadmap" for the agency is bang on schedule.

"The roadmap has really taken hold," both inside and outside the NIH, Zerhouni said at a briefing held on 14 October to mark the anniversary. "We have made a lot of progress," Zerhouni said. He added that the biomedical agency's institutes and centres have responded to his \$2.2-billion, six-year blueprint with "enormous and amazing speed".

Outside observers have endorsed Zerhouni's upbeat assessment of the roadmap's progress. Grandiose plans in government often "seem to take forever", notes David Moore, a lobbyist for the Association of American Medical Colleges (AAMC). "But they have gone in a relatively short time from the conceptual to actually having programmes in place."

At the briefing, Zerhouni provided a list of 192 roadmap project awards that have been made in 2004. But some observers are unsure that the project will outlast Zerhouni's tenure. A radiologist, Zerhouni is keen to align the NIH more closely with patients' immediate needs for medicines and technologies. This is a significant turnabout for an agency whose agenda has traditionally been driven by basic research in molecular biology.

But Zerhouni said that extramural researchers have met "every single" roadmap initiative with a more enthusiastic response than the NIH had expected. Some 1,000 researchers applied for nine extramural Pioneer Awards, which provide \$500,000 of funding each year for five years, with no strings attached.

The roadmap aims to get NIH institutes and centres to work towards three broad goals: better systems-biology tools; enhanced interdisciplinary team research; and accelerated clinical research to get discoveries from bench to bedside more quickly. The specifics include a network of biocomputing centres, new interdisciplinary research centres and innovative training for clinical researchers (see *Nature* 425, 438, 545; 2003).

The whole project is consuming less than 1% of the NIH's \$27.8-billion budget, including \$129 million in 2004. "But it is an important percentage," Zerhouni said last week, "because it's what we call 'risk money'."

Moore says that the plan is being supported



On target: Elias Zerhouni (right) and institute heads.

in the Congress. The feeling among legislative staffers is that it "is giving a direction to the NIH", he says. "It's not just more of the same."

Anthony Means chairs the department of pharmacology and cancer biology at Duke University Medical Center in Durham, North Carolina, where researchers have won seven roadmap grants this year. Means says that, as president of the Endocrine Society, he backs the roadmap because of its emphasis on getting basic science discoveries to patients. For the first time in his 30 years in science, Means says, he's hearing an NIH director clearly say: "If you expect to be funded by the NIH, you're going to have to be willing to consider the application of the science that you're doing."

That message has met with mixed reviews from investigators, Means adds: "Some people have taken to it pretty well. And others say they don't want to be told how to do science."

Still others are sceptical that, in an era of essentially flat NIH funding, the roadmap can outlive Zerhouni's tenure. "There is some fear that it won't be a sustainable thing," says Tony Mazzaschi, associate vice-president for research at the AAMC. Mazzaschi points out that the initiative's funding is predicated on voluntary yearly contributions from all 27 NIH institutes and centres. When funding gets tight, and the choice is between funding a roadmap initiative and an institute's own best research grant proposals from individual investigators, Mazzaschi says, "some of these cross-institute initiatives may be a lesser priority for a given institute. As they say, 'all politics is local'." ■

Russians fear privatization of state labs

Bryon MacWilliams, Moscow

Senior officials in Moscow are attempting to reassure Russian scientists about details of a planned reform, under which many hundreds of the country's state research institutes could lose their government funding.

According to a document leaked to the Russian Academy of Sciences last month, the government is focusing on the 450 institutes run by the academy, as well as nearly 2,000 other laboratories run by the government. Russian scientists fear that tens of thousands of them could lose their jobs in the restructuring that is being proposed.

At a hurriedly arranged press conference in Moscow on 18 October, Andrei Fursenko, the minister of education and science, confirmed that the government planned to streamline Russian science. But he said that the details had yet to be determined. "The concept is not a strict directive," he said.

Fursenko added that the ministry and the academy will together determine the criteria by which institutes will be preserved, consolidated, privatized, turned into foundations or simply liquidated. Yury Osipov, president of the Russian Academy of Sciences, said that he will work with the ministry.

The academy was an important scientific powerhouse in Soviet times, although most of its strengths were concentrated in a few disciplines, such as mathematics and weapons science. But it has been criticized in



Stark choices: the Russian Academy of Sciences may face major restructuring of its institutes.

recent years for contributing little to industrial innovative power, and for underperformance in basic research.

The need to free up money for applied research has been cited as a driving force behind the planned restructuring. But worried scientists and representatives of trade unions say that they think another factor is at work: the government's desire to generate cash by privatizing land, buildings and patents.

Scientists fear that the restructuring of the academy may provide a pretext for allowing private investors to get their hands on property in one of the world's most expensive real-estate markets.

"We agree that Russian science does need reform," says Boris Stern, a particle physicist at the academy's Institute for Nuclear Research in Moscow. "But we also know that privatization of state property in Russia rarely has the desired effect. It is usually just about selling real estate to businessmen."

Irina Belyaeva, deputy director of the academy's library in St Petersburg, made a similar charge at a press conference on 30 September. "Does the state have the right to privatize the personal property of Peter the Great and his associates?" she asked. She conceded that the library might be unprofitable for the state, but added: "The value of its stock cannot be measured in money."

At Monday's press conference, there was also disagreement over future governance of the academy itself. Fursenko indicated that he views the academy as a state institution in charge of state property, whose leadership should be appointed by, and answerable to, the Russian president, Vladimir Putin. But Osipov said the academy should be only partially beholden to the state and free from "the thumb of bureaucracy".

As *Nature* went to press, trade unions and state laboratories were planning a protest in Moscow on 20 October.

Additional reporting by Quirin Schiermeier in Munich.

French delegation strengthens bond with China

David Cyranoski, Tokyo

France and China are a lot closer — economically, politically and scientifically — following the French president Jacques Chirac's visit to China earlier this month.

Travelling through the country with a 50-strong French trade delegation, Chirac helped to secure billions of dollars' worth of industrial contracts for France at the same time as signing framework agreements for the promotion of research into environmental protection, the development and peaceful use of atomic energy, health and medical science, nuclear fusion, and the exploration of space.

How such plans will be implemented is not clear, but Chirac's reported desire to "deepen our global partnership with China" is definite.

The new Institut Pasteur of Shanghai, a collaboration between the Paris-based

Institut Pasteur and the Chinese Academy of Sciences, is one of the cornerstones of this partnership. On 11 October, Chirac attended the institute's inauguration with Vincent Deubel, its recently appointed director. Deubel, who is French, is the first foreigner to head a scientific research institute on the Chinese mainland.

The institute, which is expected to open sometime this year, will focus on research into diseases that have devastated parts of China, including AIDS and hepatitis C. It is expected eventually to support 500 scientists, and to be at the centre of China's attempts to deal with its expanding problems with infectious diseases. It should bring top technology and equipment to the country, along with expertise in working in biosafety labs. "Courses on biosafety will be our first duty," says Deubel.

Other partnerships may face a more

difficult political future. During Chirac's trip, France and China agreed in principle to work together on space research. This may place France at odds with the United States, which has long blocked Chinese participation in the International Space Station.

During his visit to Shanghai, Chirac told students at Tongji University that France would increase support for Chinese students, saying that the current 8,000 students who come to France from China each year are "too few". This move comes at a time when many Chinese students are prevented from going to the United States by its strict immigration policies.

For more news and analysis go to
news @ nature.com
www.nature.com/news

Schwarzenegger endorses stem-cell push

Jonathan Knight, San Francisco

After months of silence, California's Republican governor, Arnold Schwarzenegger, lent his backing on Monday to the state's ballot initiative to fund embryonic stem-cell research. Supporters of the plan, on which Californians will vote in two weeks, had become increasingly optimistic that it would pass following a string of endorsements from newspapers and public figures, as well as a continuing lead in the polls. The governor's approval now adds to their confidence.

The measure, known as Proposition 71, would pump \$300 million a year for a decade into the ethically charged research. Its supporters say the investment would pay large dividends both in treatments for intractable diseases and in healthcare savings for the cash-strapped state. Opponents call those claims speculative at best, and point to the ethical difficulties of taking cells from embryos that are either discards of fertility treatments or created specifically for cell harvesting.

The Proposition 71 campaign turned up the heat on 24 September with a string of 30-second television spots in which researchers, patients and celebrities such as the actor Michael J. Fox, who has Parkinson's disease, urged Californians to vote 'yes'. "It could save the life of someone you love," says Fox in his advert, which first aired on 14 October.

The latest Field Poll, taken just after the ads began, showed 46% of Californians for



Governor Arnold Schwarzenegger has lent his weight to California's stem-cell initiative.

the initiative and 39% against, indicating a slight increase in its popularity since August.

Proponents have lined up a long list of supporters, including dozens of patient advocacy organizations, business groups, medical associations and Nobel laureates. Several of the state's major newspapers have also lent their support, including the *Los Angeles Times*, the *San Francisco Chronicle* and *The Mercury News*. The *Sacramento Bee* is opposed.

Although initial opposition to the measure

came largely from conservatives who object to embryo research on moral grounds, a number of liberal groups are now lining up against it on the grounds that it lacks adequate ethical safeguards against financial conflicts of interest (see *Nature* 431, 232; 2004).

"Stem-cell research needs controls and oversight. People in the progressive community have been saying that for a long time," says Susan Fogel, campaign coordinator for the ProChoice Alliance Against Proposition 71. The group's supporters include the National Women's Health Network and the California Nurses Association.

Objectors sparred with proponents before a 300-strong audience at a symposium in San Francisco on 12 October. The symposium was sponsored by former city mayor Willie Brown's Institute of Politics and Public Service. State senator Tom McClintock, a southern California Republican, said he was opposed on fiscal grounds. "I do not understand why it is suddenly the responsibility of California, with the lowest credit rating in the nation, to fund this research," he said.

But former secretary of state George Shultz, an influential Republican who now lives in San Francisco, countered that the state would benefit from the spending. "Basic research is very productive," he said. Shultz added that it would encourage businesses and researchers elsewhere to "come to California, where the action is". ■

Germany balks at funding ESA's planetary ambitions

Alison Abbott, Munich

Aurora, the European Space Agency's planetary exploration programme, may find itself in difficulties as France and Germany proved reluctant to support it last week.

Representatives of the agency's member states, meeting in Paris, heard that it had obtained commitments of between €25 million (US\$31 million) and €30 million. This is about half what the agency thought it would need to keep the programme going until it starts in earnest in 2006.

The shortfall is due to notable absences among the countries that have pitched in: Germany, in particular, has decided to withhold funding completely.

Europe is winding down its contribution to the International Space Station, and will instead focus on Aurora. The aim for the programme is to ensure that Europe maintains expertise in cutting-edge space technologies, so it can be an equal partner in any future international planetary exploration projects.

Launched in 2001, Aurora's preliminary

projects are being updated in line with recent findings. Its long-term continuation must be approved by the agency's ministerial council at the end of next year.

Germany and France are the largest contributors to the European Space Agency (ESA). Together they account for more than half of its total budget: Germany paid 41% of Europe's bill for the International Space Station, for example. Costs for that project have hugely overrun, and Germany is still paying out. Together with France, Germany has also had to provide much of the increasingly hefty subsidy for Ariane launchers — the boosters that provide Europe with independent access to space. This independence from other agencies is fundamental to ESA policy.

Against this backdrop, Germany has decided to contribute nothing to Aurora. France has pledged a maximum of only €2 million and may also decide to give nothing when it makes its final decisions on financing at the end of this year. This puts Aurora in the uncomfortable

position of being unsupported by ESA's two main players.

Surprisingly, Italy — the third largest contributor to ESA — has taken the lead and offered Aurora up to €14 million (see *Nature* 431, 619; 2004).

"We would like Germany to be able to join eventually, and hope for a higher contribution from France," says Aurora's director Daniel Sacotte, based at ESA's research and technology centre in Noordwijk, the Netherlands.

There are a few years yet for countries to change their minds. The first Aurora mission — a rover called ExoMars — is expected to launch no earlier than 2011. To fund this and other projects, Sacotte hopes for an annual budget of €200 million. But this could be extremely optimistic if major contributors do not jump back on board.

Scientists at Germany's space agency, the DLR, say that it is unlikely the country will be able to contribute to Aurora any time soon, as problems with the International Space Station continue. ■

Quarrel over book leads to call for misconduct inquiry

Rex Dalton, San Diego

A political scientist and former journal editor has asked the US National Science Foundation (NSF) to set up an inquiry to determine whether scientific misconduct was involved in the writing of a psychological history book.

Gary Johnson, chair of the political science department at Lake Superior State University, Michigan, and former editor of the journal *Politics and the*

Life Sciences, wrote to the inspector-general's office at the NSF on 27 September, requesting a probe into the book *Born to Rebel: Birth Order, Family Dynamics and Creative Lives* by scientific historian Frank Sulloway.

In March this year, the journal published an edition almost entirely dedicated to criticism and commentary on the 1996 book, which argues that scientists and religious and political leaders who have older siblings are more likely to live rebellious lives (see review, *Nature* 384, 125–126; 1996). The edition of *Politics and the Life Sciences* carrying the critiques (19(2); 2000), which is not yet available online, had been delayed for four years by legal threats by Sulloway and associated problems.

Articles in the journal allege that Sulloway may have selected and manipulated data to prove his theories about how the order of birth of scientists and political leaders may have shaped their lives. "The evasiveness, the varying methodological accounts, the data discrepancies and the seemingly desperate attempts to interfere with publication together suggest that an independent review of Sulloway's research should be undertaken," says Johnson, who edited the journal for ten years until 2001, and then retained editorial responsibility for the edition on the Sulloway book.

Sulloway, a former MacArthur Fellow and a visiting scholar at the Institute of Personality and Social Research at the University of California, Berkeley, strongly denies the allegations. He charges Johnson with "unprofessional editorial tactics" during the years of disagreement over the critical journal edition. In 2000, Sulloway even threatened to file a misconduct complaint of his own against Johnson with Robert Arbuckle,



Gary Johnson requests review of data in birth-order book.

scholar at the Massachusetts Institute of Technology, Cambridge, when the book was published. Asked about the request to the NSF, Sulloway said he was unaware of it. "The only concern to me is the amount of time it takes to make all the data available," he says. "I'm just swamped with work."

Politics and the Life Sciences is published twice a year by the University of Maryland, having lost its publishing deal with Beech Tree Publishing of Guildford, UK, during the Sulloway disagreement.

The theories discussed in *Born to Rebel* about how birth order may influence personality have been hotly debated by psychologists for decades, and its high-profile publication reinvigorated that debate.

In late 1998, a critical analysis of it was submitted to *Politics and the Life Sciences* by Frederic Townsend, a Chicago-based commodity trader who had studied birth order in college. He alleged that Sulloway had selected data points

to produce supportive figures, misused study results and manipulated personality characteristics of famous studied people. Townsend says he was surprised but delighted when peer reviewers — seven out of eight, in the end — recommended publishing his 22-page article, which appeared with a 34-page editorial by Johnson. "An independent review of Sulloway's research is long overdue," Townsend says.

Sulloway says he never tried to halt the journal's publication, acknowledging that a letter to Johnson in 1999 threatening legal action may have been too "strongly worded". He points to commentaries by scientific editors and friends who support his criticism of Johnson.

As is usual in such cases, the NSF declined to comment on the inquiry request. ■

Germany's junior professors fight for their rights

Quirin Schiermeier, Munich

Guido Fischer has an identity problem. The 36-year-old microbiologist at the RWTH-Aachen University in Germany has no idea what to call himself. On his letterhead, he uses the bizarre *Vorgriffsjuniorprofessor* — which roughly translates as 'junior-professor-to-be'.

Fischer's dilemma is the result of a court ruling in July, which declared his chosen career path unconstitutional (see *Nature* 430, 599; 2004). His title of 'junior professor' was created in 2002 by a federal law that aimed to provide a fast-track to tenure for Germany's young scientists. But some of Germany's individual states, or *Länder*, took issue with the federal government's intrusion on the way they oversee science and education. Germany's highest court agreed that the federal government had no right to intrude, and some 600 young researchers were left unsure of their future.

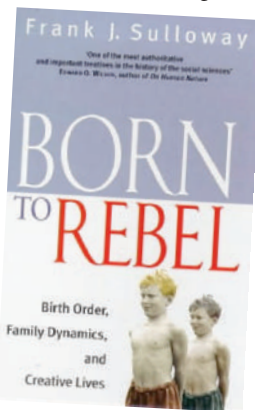
At a meeting last week in Berlin, about 150 of these 'junior professors in limbo' called on federal and state science ministers to sort out the resulting mess.

Although initially employed under the same conditions, junior professors in different parts of Germany now have quite different rights and academic status. Those employed by universities in Berlin or Hanover, for example, can still call themselves professors, whereas their colleagues in Dresden or Aachen cannot.

The *Förderverein Juniorprofessur* — a lobby group for the junior professors — last month drafted a proposal on how to restore equal status and improve career prospects. At last week's meeting, Edelgard Bulmahn, Germany's science minister, promised to include key points of their proposal in a revised federal university-education law, due to be introduced to the cabinet by the end of the month.

The *Länder* have indicated that they will not challenge the revised law, particularly as it now omits a promise to phase out the traditional method of obtaining tenure in Germany — a process called *Habilitation* that requires researchers to complete a second thesis. Some of the more conservative German *Länder* strongly opposed this change.

The revised law would reinstate the position of junior professorships, but critics say that the continued presence of *Habilitation* as a career option will make it difficult for junior professors to compete with other scientists for jobs. ■



Just cause? Frank Sulloway investigated family dynamics.



Fighting the plague: local girls try to fend off incoming locusts that threaten their crops.

African locust swarms spread north leaving trail of devastation

London Swarms of locusts that have ravaged the savannahs of western Africa are sweeping north, bringing control efforts to “a critical phase”, say officials at the United Nations’ Food and Agriculture Organization (FAO). The locusts have reached southern Libya and Algeria, and the borders of Morocco.

The destruction is “massive”, says FAO

director-general, Jacques Diouf. Officials say the plague is the worst for 15 years, but they will not know the full extent of the damage until investigators report back in November. The locusts have devastated the livelihoods of farmers, shepherds and even beekeepers, whose bees are killed by pesticides used to fight the locusts.

A funding drive has so far raised more than US\$40 million in aid, some of which will go towards orders for vast quantities of pesticide. But FAO officials say that much more money is needed.

Germany adds its plank to the open-access platform

Munich The German government has shown its support for the open-access movement by giving the Max Planck Society €6.1 million (US\$7.6 million) to develop an access system for publications from all of the society’s 80 or so institutes.

The society is planning a six-month study to investigate how scientists would like such a system to work. It will then collaborate with the scientific and technical information services institute FIZ Karlsruhe to develop a database that will serve the general public as well as researchers. The government would like the resulting system, called MPS eSciDoc,

to serve as a model for other research organizations, it says.

The Max Planck Society was one of the key signatories to the Berlin Declaration in October last year, which pledged to ensure free access to the results of all publicly funded research through the Internet.

Malaria trial boosts hopes for a vaccine

Washington A vaccine has successfully protected some children from malaria in an African clinical trial. The results are the most promising so far for such a vaccine, which researchers have pursued for decades.

The vaccine, called RTS,S/ASO2A, was tested on 2,000 children aged between one and four in Mozambique. Those given the jab were 30% less likely to show fever or early signs of the disease after six months, and nearly 60% less likely to have symptoms of more severe disease (P. L. Alonso *et al.* *Lancet* **364**, 1411–1420; 2004). The vaccine also protects adults from malaria, but only for a few weeks. It seems to offer more hope for children, who suffer most deaths from malaria.

Researchers hope to combine RTS,S/ASO2A, which fights the malaria parasite at an early stage in its life cycle, with vaccines that attack later stages.

Genesis crash blamed on designers flipping a switch

Washington Investigators may have found out what caused the Genesis spacecraft to crash into the Utah desert this September: some crucial switches were installed backwards.

The craft's design drawings indicated that some tiny cylindrical plungers, intended to detect entry into the atmosphere and deploy a parachute, should be installed the wrong way round. So they were.

Investigators had to use X-rays of the wreckage to confirm that all four such gravity switches were put in backwards, including the back-ups, which explains why the parachute failed to open (see *Nature* 431, 234; 2004). None of NASA's review processes picked up the \$260-million mistake.

The Stardust mission — set to return to Utah in 2006 with samples from a comet's tail — has the same switches, but its diagrams indicate that they were installed correctly.

The Genesis investigation is expected to conclude in November.

Election puts agencies into funding doldrums

Washington The US Congress has left the budgets of some major research agencies pegged at last year's spending levels until

Croaking frogs provoke a chorus of concern

Washington The world's frogs, newts and toads are dying. They are being over-harvested for food, their homes are being destroyed, and entire species are disappearing for no apparent reason.

That is the conclusion of more than 500 herpetologists worldwide. They have known for some time that many amphibian species are in trouble, but their report (S. Stuart *et al.* *Science* doi:10.1126/science.1103538; 2004) is the first global assessment of the creatures' status.

Simon Stuart of the World Conservation Union (IUCN) in Washington and colleagues examined the status of all 5,743 known species. They found that more than 30% qualify as vulnerable, endangered or critically endangered — including such colourful species as the harlequin toad (right), whose numbers have crashed dramatically.



Stuart says there is a growing consensus that both climate change and a fungal disease are to blame for the decline. "Because of their sensitivity, amphibians are the first species we would expect to show adverse reactions to climate change and emerging diseases," he says.

after next month's election on 2 November.

Senators and representatives failed to pass 10 of 13 bills before heading home for a final round of campaigning for the election.

Their inability to reach a consensus — which was widely anticipated earlier this year (see *Nature* 430, 388; 2004) — will leave federal research agencies funded at fiscal year 2004 levels until well into the

2005 fiscal year, which began on 1 October.

The delay means that some agencies are uncertain about their future funding. NASA, for instance, faces a \$1.3-billion gap between the House bill, which favoured a cut to \$15.1 billion, and the Senate version, worth \$16.4 billion. The matter is now expected to remain unresolved until the new Congress convenes in January.

A shot in the arm

Antibiotics are failing and drug companies have all but stopped developing new ones. Will conquered diseases come back to haunt us? Martin Leeb examines one plan to avert the crisis.

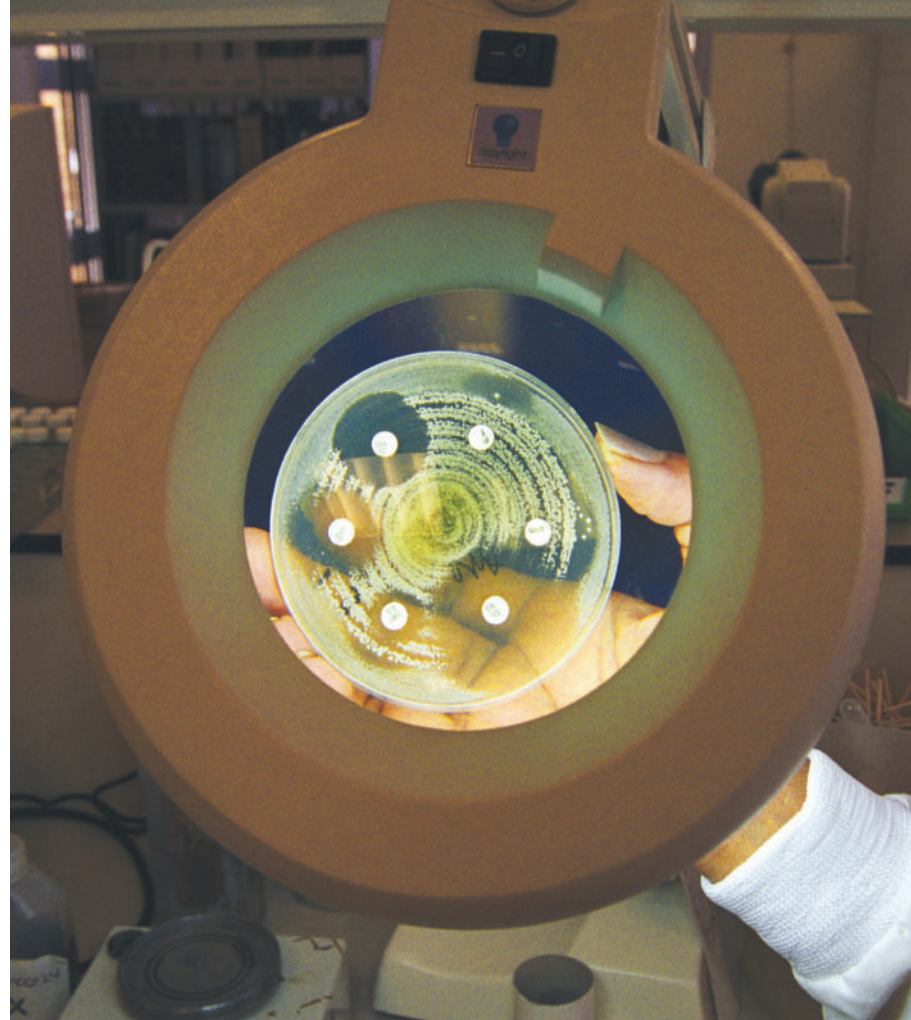
In the late 1960s the battle against bacterial infections was considered won, in the developed world at least. By the time of Woodstock, antibiotics were curing previously lethal infections in a matter of days. Infected cuts and food poisoning were no longer life-threatening, diseases such as syphilis and gonorrhoea seemed to be on the way to eradication, and ancient scourges such as plague and cholera could now be controlled.

Now antimicrobial resistance threatens to turn back the clock. Resistance is spreading rapidly, particularly in hospitals, where many different bacterial strains can come into contact with each other and where antibiotics are heavily used (see graph, page 893). The catch is that the more an antibiotic is used, the more resistance to it spreads, forcing doctors to try other antibiotics. Even drugs that once served as a last resort are losing their potency.

"The cycle of resistance is inevitable," says Christopher Walsh, a molecular pharmacologist at Harvard Medical School in Boston. "With every new drug we use, we select for resistant microbes that survive and multiply unhampered by treatment with the same antibiotic."

The only solution, experts agree, is to continue to develop new drugs. But just as we need them most, the antimicrobial drug pipeline is running dry. Until ten years ago, all major drug companies ran antibacterial research programmes. Today, these programmes have been drastically pruned, and many have been cut altogether as companies pursue more lucrative areas, such as chronic illnesses and mood disorders.

The desperate nature of the situation led the Infectious Diseases Society of America (IDSA) to issue a white paper in July calling for a variety of measures to get antibiotic research



Zero impact: bacterial growth around the white antibiotic discs spells a resistant strain.

back on track, starting in the United States. The report — called *Bad Bugs, No Drugs* — followed a year-long investigation into the economics of drug development. In the absence of independent action by the pharmaceutical industry, the report says, the US Congress and federal regulatory agencies must step in with financial incentives for companies to get back into the antimicrobial business.

Last resort

This all makes for a potential healthcare calamity. Although the number of hospital-acquired infections has been gradually declining in the United States, a greater proportion of them — now about 70% — are resistant to at least one antibiotic. This results in a delay in effective treatment, prolonging illness and increasing the risk of death. Two million people will pick up an infection in a US hospital this year, for example, and 90,000 of them will die of it, according to estimates by the Centers for Disease Control and Prevention in Atlanta, Georgia.

Because bacteria share resistance genes, antimicrobial resistance fostered in hospitals ultimately spreads to surrounding communities. As a result, hard-to-treat forms of old killers such as tuberculosis and cholera are showing up in many parts of the world.

Even vancomycin — still used only as a last resort because of its side effects — is losing its punch. Vancomycin-resistant gut bacteria first showed up in 1986. Then in 1997, the discovery of partially resistant strains of *Staphylococcus aureus*, which causes serious wound and surgical infections, shattered the hope of maintaining vancomycin as the ultimate weapon against the worst hospital-acquired infections. Two years ago, the first cases of fully vancomycin-resistant *S. aureus* were reported in the United States.

"This should make you very pessimistic," says Walsh. "If they can become resistant to vancomycin, they will become resistant to everything."

Most initiatives to deal with the situation have concentrated on delaying the inevitable build-up of resistance by reducing the indiscriminate use of antibiotics in medicine and agriculture. Several European countries, for instance, have phased out the use of antibiotics as growth promoters in food animals since the mid-1990s, resulting in reduced rates of resistance¹.

But such measures only slow the spread of resistance. Despite the pressing need for new antibacterials to replace those that are losing their effectiveness, pharmaceutical companies are not beefing up their research efforts in this field. Why?

They aren't for the very same reason that

"If they can become resistant to vancomycin, they will become resistant to everything"

— Christopher Walsh



A cocktail of antibiotics is used to treat a patient with septicaemia (above, top), a bloodstream infection. *Staphylococcus aureus* (above) has become resistant to almost all antibiotics.

the problem is so intractable. Microbial resistance increases the demand for new products but at the same time it shortens their useful life, severely impairing a drug's long-term potential to return a profit. And as the best antibiotics are often held in reserve, a new antibiotic to which there is no resistance would be little used. That is not the sort of product a profit-making company wants to develop.

Furthermore, from a marketing standpoint, antibiotics are the worst sort of pharmaceutical because they cure the disease. Companies have more incentive to bankroll research into treatments for chronic conditions such as high cholesterol or rheumatoid arthritis, for which patients take the drug over years or a lifetime,

rather than for just a week or two.

The cost of drug development adds to the pressure for high-return products. According to the Tufts Center for the Study of Drug Development in Boston, it costs some \$800 million on average over 10–15 years to bring a new drug to market². This cost, combined with a likely low return on investment, makes pharmaceutical companies reluctant to pursue antibiotic research.

Patent holiday

The result of these negative pressures has been a 56% decline in the number of antibiotics approved annually by the US Food and Drug Administration (FDA) over two decades (see the Commentary on page 899 of this issue). According to a study published in May this year, out of 506 drugs in late-stage clinical testing by the world's 15 largest pharmaceutical companies, only six are new antibacterials — and all are derivatives of known antibiotics³. That is barely more than the number being developed for some less life-threatening ailments. Erectile dysfunction alone accounted for four of the new drugs being tested.

“Unless things change soon, we are going to face a major health crisis,” warns Brad Spellberg, an infectious-disease researcher at Harbor-UCLA Medical Center in Los Angeles and lead author on the study.

To help avert the crisis, the IDSA is recommending several steps that would radically alter the economic landscape for drug companies. First, it has proposed that Congress establish an independent commission to prioritize antimicrobial discovery, which would identify the biggest microbial threats and make recommendations directly to the Secretary of Health. Pharmaceutical companies could then register with the Department of Health and Human Services to receive tax credits if they were working on drugs to fight one of the targeted bugs.

Another carrot would be ‘wild-card’ patent extensions awarded to any drug company that successfully developed a new antibiotic to a targeted microbe. This would allow the company to extend the life of any one of its patents — not necessarily an antibiotic — for two years. As some blockbuster drugs bring in billions of dollars a year, that amounts to quite a gift. “I think that patent extension will increase profits substantially,” says Joseph DiMasi, director of economic analysis at the Tufts Center.

At the same time, the FDA should set new drug-approval procedures for antimicrobials, the IDSA says. At the moment, a new antimicrobial must be proved more effective than an existing one against drug-sensitive bacteria. But this is nearly impossible, as current antibiotics are very effective as long as the bacteria are not resistant. The alternative,

testing in patients with drug-resistant infections, is much more expensive, because it is difficult to find enough patients for a complete trial.

The way round this is to extrapolate from tests against drug-sensitive strains to gain insight into how effective the antibiotic will be against drug-resistant strains. But the FDA has no standards for allowing this, much to the disappointment of drug companies says Steven Projan, director of antibacterial research at Wyeth in Pearl River, New Jersey. “We need a roadmap for getting a drug approved,” he says.

Additional incentives proposed by the IDSA include extending patents to make up for time lost during FDA review, establishing a guaranteed market with federal dollars and limiting liability for adverse effects, as has been done for vaccines. All these measures would require legislation by Congress.

Convincing Congress to act may be difficult, as congressmen have a proven distaste for wealthy pharmaceutical companies complaining about their bottom line. “The poor image of pharmaceutical companies with Congress is a potential obstacle,” says DiMasi.

In fact, some of the provisions were introduced in the Senate in 2001 and 2003 as part of a bill known as S. 666, written by Senators Joe Lieberman (Democrat, Connecticut) and Orrin Hatch (Republican, Utah). But S. 666 never came to a vote. Instead, parts were retained in what became Project BioShield, the bill signed into law in July this year by President Bush to fund stockpiling of vaccines and drugs against bioterror agents such as smallpox and anthrax.

But Project BioShield does nothing to protect the public against the demise of antibiotics. Now Lieberman and Hatch are planning to introduce Project BioShield II later this year, a bill that will combine features of both S. 666 and the IDSA recommendations.

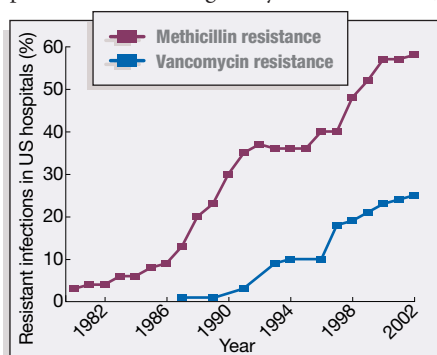
Even if the IDSA plan is adopted, there are still considerable scientific hurdles to clear. Nearly all the existing classes of antibiotic are half a century old. Most antibiotics are merely modifications of existing ones, which does not always solve the problem of resistance. Since 1962, only two new classes — an oxazolidinone (linezolid) in 2000 and a cyclic lipopeptide (daptomycin) in 2003 — have been approved for use.

“We have already collected the low-hanging fruit,” explains Projan. But in every orchard the tastiest fruits hang the highest. The proposed incentives, say Projan and others, may provide a ladder to reach them. ■

Martin Leeb was until recently an intern in the Munich office of Nature.

1. Wegener, H. C. *Curr. Opin. Microbiol.* **6**, 439–445 (2003).
2. DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. *J. Health Econ.* **22**, 151–185 (2003).
3. Spellberg, B., Powers, J. H., Brass, E. P., Miller, L. G. & Edwards, J. E. *Jr Clin. Infect. Dis.* **38**, 1279–1286 (2004).

SOURCE:



DEAF by design

Employing genetic diagnosis to avoid having a baby with a disability is controversial enough. But a minority of deaf people would consider testing to ensure that they had a deaf child. Carina Dennis finds out why.

John and Karen — not their real names — are both deaf, and desperately wanted a deaf baby. But genetic testing showed that this was extremely unlikely. “They were devastated,” recalls Arti Pandya, a clinical geneticist at Virginia Commonwealth University in Richmond, who counselled the couple. It was two years before they got over their disappointment and started trying to conceive their first child.

The couple’s attitude will shock many people. If you can hear, it’s hard to understand why anyone would want a deaf child. But John and Karen’s views are not that unusual among those who identify themselves as ‘Deaf’ with a capital ‘D’. The Deaf view their condition not as a disability, but rather as the underpinning of a rich culture that should be celebrated and preserved. And with the identification of the most common genetic mutations linked to deafness, it is now possible, in theory, to make an active choice to have a deaf child.

This possibility turns the debate over designer babies on its head, providing ethicists and genetic counsellors with a dilemma. Only a tiny minority of deaf people would wish to use genetic tests in this way. Some argue that their reproductive choices should be respected. But is society prepared to sanction the use of genetic diagnosis for a purpose that many find difficult to understand — and some might even see as immoral?

Some Deaf people despair of ever being understood by those who aren’t part of their culture. The Deaf identity is in large part a product of a shared sense of isolation from the hearing world. “Exclusion is central to



All together now: deaf culture now encompasses everything from spelling bees (audience shown applauding, above) to Broadway shows (right).

the experience,” says Gary Kerridge, regional disability liaison officer at the University of Ballarat in Mount Helen, Australia, who lost his hearing as a young child.

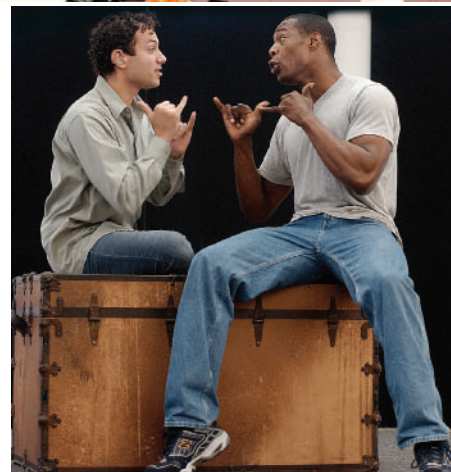
For deaf children, the majority of whom are born to hearing parents, even family gatherings can be lonely affairs. Many of them feel liberated by their first experience of Deaf culture. “They learn to sign and suddenly for the first time, after years of being isolated and struggling, they are accepted,” says Kerridge. “Naturally, they quickly develop a strong attachment to the Deaf way of life.”

A world of their own

Sign language is central to the lifestyle. It uses hand shape, position and movement, plus posture, facial expressions and other visual cues, to form words and convey meaning. It has its own rules for grammar, punctuation and sentence order. It is elaborate and expressive, and lends itself readily to poetry and theatre.

For a hearing person, entering a room full of chattering signers can be disconcerting. Methods used to attract attention, for example, seem downright rude. “Stomping on floors, waving animatedly, flashing lights and thumping tables are all considered OK,” says Kerridge.

Knowing sign language doesn’t, by itself, break down the barriers between the hearing and the Deaf. “Even hearing people from Deaf families and who sign well are always, to



G. OMENICO/AP

a certain degree, seen as culturally distinct,” says Kerridge. “That absolute feeling of exclusion from the hearing world is difficult for a hearing person to fathom.”

Within Deaf culture, however, there’s a level of social intimacy that is rare among the hearing. “I will meet another Deaf person for the first time and in five or ten minutes, it’s not uncommon to know a great deal about their family and personal life,” says Carol Padden, a linguist at the University of California, San Diego, who was born deaf, to deaf parents. “I have to remind myself not to expect the same invitation to become familiar when I’m with hearing colleagues.”

That, in a nutshell, is why some deaf couples would prefer to have deaf children. Communication and the pursuit of intimacy are central to being human. If you genuinely believe that your children will have at least as rich an emotional life if they cannot hear, and



Good vibrations: the deaf community's experiences, such as 'listening' to a concert through a balloon, can be difficult, if not impossible, to explain to the hearing.

that you will be better able to communicate with them, why not make this choice?

"I don't see anything wrong with it. I see it as being similar to how parents determine the religion or education of their child," says Ted Supalla, who has been deaf since birth, and studies sign languages at the University of Rochester in upstate New York. Supalla's own children can hear; they communicate with him by sign language and speak to his hearing wife.

Genetic lottery

Like Supalla, most deaf people are happy to let nature take its course, and say that they would be content to have a hearing child. But deaf people are increasingly marrying one another, making deaf children more likely. A report published in April theorized that the increasing number of marriages among the deaf during the nineteenth century may have doubled the frequency of deafness in the United States caused by mutations in genes for proteins called connexin 26 and connexin 30, which affect the function of the ear's sound-sensitive cochlea¹.

About 1 in 1,000 infants is born profoundly deaf. About half of these cases have a genetic cause. Mutations in many genes are involved — the most common, accounting for about one in five deaf children, are those affecting connexin 26.

Still, most children born to deaf couples can hear. Many of the mutations involved are recessive, which means that a baby will be

deaf only if it inherits two copies of the same mutated gene. For John and Karen, the laws of inheritance could not give them a deaf child — their deafness is due to recessive mutations in different genes.

The couple's genetic counsellor is now investigating attitudes to genetic testing among the deaf. In a pilot study conducted at Gallaudet University in Washington DC, a college for the deaf and hard-of-hearing, Pandya and her colleagues asked students whether they would be interested in considering genetic test results to help them select a partner². More than half of the 64 respondents said they would — but it wasn't clear from the wording of the questionnaire whether this was because they wanted a deaf child, or a hearing one. Pandya is planning a larger study to explore the issue further.

Using genetic tests to identify a partner with whom to try and have deaf children is one thing; aborting a fetus if it turns out to be able to hear is another. Evidence that a small minority of deaf people would consider this option comes from the work of Anna Middleton, a genetic counsellor at Addenbrooke's Hospital in Cambridge, UK.

Middleton's first survey was conducted at the Deaf Nation conference, a gathering of the culturally Deaf held in Preston in north-west England in 1997. Of the 87 delegates who completed the questionnaire, 14 said they would be interested in prenatal testing for deafness. Four of these said that they would prefer to have deaf children³.

Critics argued that Middleton's study was too small, and was based on a group of Deaf activists⁴. So she polled a larger sample of the hard-of-hearing, hearing people with deaf family members, and profoundly deaf people — two-thirds of whom were not culturally Deaf. Across the deaf group, about one in five said they would consider prenatal genetic testing, mostly to prepare for the birth of a hearing or a deaf child⁵.

Few of the deaf respondents said they would consider abortion, and in most of those cases, their choice was actually for a hearing child. None of those who said they would abort a deaf fetus was culturally Deaf. But three deaf people said they would consider aborting a fetus if it could hear. Two of these were culturally Deaf.

Tough choices

Middleton says that it's still unclear what people would do when faced with the choice for real. "Attitudes do not necessarily predict behaviour," she cautions. And even among Deaf activists, it's hard to find someone who will be quoted as saying they would abort a hearing fetus, because of the opprobrium they would attract. "Deaf people know that it's a very risky thing to say in public that you would consider genetic testing to have a deaf child," says Padden.

The wisdom of keeping quiet was reinforced by the controversy that engulfed Sharon Duchesneau and Candace McCullough in April 2002. A Deaf lesbian couple from Bethesda, Maryland, Duchesneau and McCullough told the *Washington Post Magazine* that they had conceived a child using sperm donated by a deaf male friend, because they wanted a deaf baby. They didn't employ genetic testing to guarantee success, but their son, Gauvin, was born deaf. While the initial article was sympathetic, many of those that followed were not. The Fox News website, for instance, ran a hostile piece, headlined "Victims from birth: engineering defects in helpless children crosses the line".

Deaf couples wanting to be sure of having a deaf child have two options. They could use prenatal genetic testing, and abort the fetus if it can hear. Or they could consider *in vitro* fertilization (IVF) combined with preimplantation genetic diagnosis to select deaf embryos for transfer to the womb. In December 2002, Monash IVF, a clinic in Melbourne, Australia, conducted preimplantation tests for a couple who wanted to exclude the one-in-four chance that they would have a deaf baby.

The Infertility Treatment Authority for the state of Victoria, which sanctioned the Monash procedure, says it would not allow a couple hoping for a deaf child to use the test. "Our policy states that the procedure should be used to avoid a genetic abnormality," says Helen Szoke, the authority's chief executive. Few other regulatory bodies have yet devised



Handmade: the Café Signes in Paris is designed to bring locals and the deaf community together.

explicit policies on the issue. Britain's Human Fertilisation and Embryology Authority, for instance, which issues licences for preimplantation genetic testing on a case-by-case basis, has not yet had to rule on the matter. The UK Human Genetics Commission, meanwhile, is currently preparing a report for the government on genetics and reproductive decision-making, which may touch upon the issue.

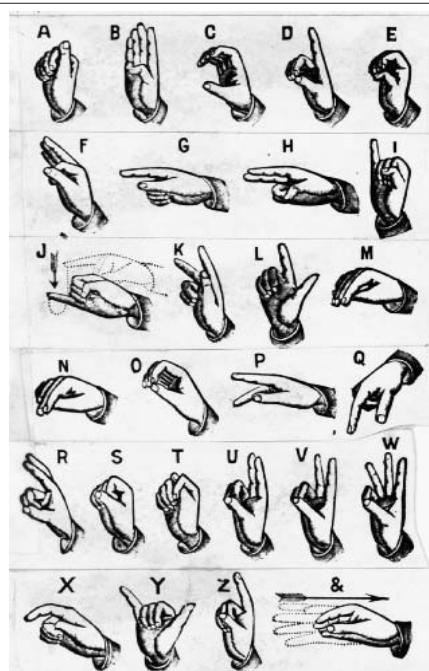
Prenatal genetic testing for hereditary conditions is used more widely than preimplantation diagnosis. And in many countries, including the United States, there are no legal restrictions on its use. Instead, clinical geneticists and genetic counsellors would have to decide whether to assist a deaf couple to have a deaf child by giving them a test that could lead the parents to abort a hearing fetus.

An international survey of 2,906 geneticists in 36 nations revealed varying views on this point. In Norway, none of those surveyed would perform such a test, and in France, the figure was just 1%. But in the United States, Italy, Russia, Cuba and Israel, more than a third said they would⁶.

In practice, such tests are far more likely to be used by hearing couples to avoid having a deaf baby. In July, *The New York Times* highlighted the case of a couple who had taken a series of genetic tests before conceiving to be sure that they weren't at risk of passing on a genetic disease. When their deaf son was born, the parents were angry that they hadn't been tested for the common



Signing is an intimate form of communication in some families.



Signs of the times: an early alphabet for the deaf.

mutations that can cause deafness.

For many people born deaf, including Padden, the attitudes revealed in the piece struck close to home. "That article sent chills down my spine," she says. Middleton's surveys suggest that many deaf people feel similarly. The culturally Deaf, in particular, feel threatened by the possibility of genetic diagnosis leading to the abortion of deaf fetuses³. Some postings on deaf online forums have equated genetic testing with Nazi-style eugenics. Similar attitudes underpin widespread Deaf opposition to the idea of 'curing' deaf people using cochlear implants.

Testing times

This unease may explain why Middleton's surveys have shown that deaf people are less likely than the hearing to consider prenatal testing for deafness^{3,5}. And among those who would consider testing, opinions vary widely. Many deaf people, for instance, are appalled by the idea of aborting a fetus if it can hear. Opinions may depend in part on whether the individual was born deaf or lost their hearing later on, and whether they grew up in a deaf family.

Given these diverse viewpoints, some experts argue that it's unfair to focus on the minority of the culturally Deaf who say they would consider aborting a hearing fetus. "It is offensive to keep harping on about this scenario. While many deaf parents may harbour a preference for having deaf children, the data suggest that the majority would never consider doing it," says Barbara Biesecker, a genetic counsellor at the National Human Genome Research Institute in Bethesda.

But if genetic testing to screen against deafness takes off, and the Deaf feel that their culture is threatened, it's possible that some will want to fight back. In this case, their best option might be to adopt the very technology they fear, and embrace genetic testing to ensure that they have deaf children.

It's even possible that some may have already done so, without anyone realizing. In many countries, there are no legal obstacles to stop a woman obtaining a prenatal test for deafness, without revealing her true motivations, and then seeking an abortion from a different healthcare provider if the result showed that she was carrying a hearing fetus. "If the question is whether there are any restraints to prevent somebody from doing this, the answer is no," says Biesecker.

Carina Dennis is Nature's Australasian correspondent.

1. Nance, W. E. & Kearsey, M. J. *Am. J. Hum. Genet.* **74**, 1081–1087 (2004).
2. Taney, P. R., Pandya, A., Foley, D. L., Nicely, L. V. & Arnos, K. S. *Am. J. Med. Genet.* **130A**, 17–21 (2004).
3. Middleton, A., Hewison, J. & Mueller, R. F. *Am. J. Hum. Genet.* **63**, 1175–1189 (1998).
4. Michie, S. & Marteau, T. M. *Am. J. Hum. Genet.* **65**, 1204–1208 (1999).
5. Middleton, A., Hewison, J. & Mueller, R. J. *Genet. Couns.* **10**, 121–131 (2001).
6. Wertz, D. C., Fletcher, J. C., Nippert, I., Wolff, G. & Ayme, S. *Am. J. Bioeth.* **2**, W21 (2002).

The real dirty secret of academic publishing

You can get almost anything into print if you go far enough down the ranks of journals.

Sir—I cannot in all honesty share in the anxiety surrounding publication of a dubious paper on ‘intelligent design’—regarded by most scientists as a version of creationism—in a journal with an impact factor of less than one. Your News story “Peer-reviewed paper defends theory of intelligent design” (*Nature* **431**, 114; 2004) suggests that getting an intelligent-design paper into a peer-reviewed journal is a huge achievement for creationism. I am more surprised it took so long to get one in.

The paper in question presents no new arguments and is unremarkable in any way except in that it has been published. It appeared in a journal that, until this particular editorial decision, enjoyed

much-deserved obscurity. Proponents of intelligent design would have us believe that this publication is a testament to the scientific legitimacy of their theory—although the editor has since left and the journal has disowned the paper as “inappropriate” (see *Nature* **431**, 237; 2004).

In my opinion it is yet another testament to the rampant proliferation of scientific publications, resulting in a flood of inconsequential papers appearing in those thousands of journals that exist on the fringes of scientific publication.

The editors and reviewers of many low-impact journals cannot provide the quality reviewing process one gets with *Nature*, *Science*, *Cell* and a few (very few indeed)

other established magazines, but any of them can affix the stamp of legitimacy to their outpourings by formally following the ‘peer-review’ protocol.

Let’s admit it—and this is the real dirty secret of academic publishing—one can publish just about anything if one goes far enough down the list of impact factors. There are papers all around us containing problems glaring enough to fail their authors in undergraduate midterm exams. The only reason they are not in the spotlight is because they do not deal with the theory of intelligent design.

Vladimir Svetlov

Department of Microbiology, Ohio State University,
484 W 12th Avenue, Columbus, Ohio 43210, USA

US restrictions limit anthrax networking

Sir—With reference to Thomas May’s Commentary article “Isolation is not the answer” (*Nature* **429**, 603; 2004), we would like to comment on European progress towards the goal of harmonizing international research on high-risk biological agents. Some of the obstacles along the way have been noted by J. van Aken and colleagues in Correspondence (“Biosecurity must be internationally supervised” *Nature* **431**, 17; 2004).

Activities that speed the development of viable vaccines, therapeutics and diagnostic tools are a key component to combat the threat of bioterrorism. To this end, the European Commission has mobilized start-up funds to strengthen networking activities among researchers, industry and the public-health sector.

One such network, Anthrax-EuroNet (www.anthraxeuro.net.org), unites leading anthrax researchers in France, the United Kingdom, Germany and Italy to discuss ways to harmonize research practices, to exchange information and materials, and to strengthen networking with countries in and beyond the European Union.

One goal is to develop a handbook of current and recommended protocols to improve comparison and interpretation of research data. Therefore, we circulated a questionnaire to all leading anthrax research labs working on vaccines and therapeutics. Because much research on this topic is done in the United States, the input of US scientists was essential to this survey; the feedback we received from them was supportive. The survey revealed that the possible exchange of information

was restricted by US regulations. In general, US scientists seem unsure how to interpret existing rules and fearful of releasing what could be considered sensitive information.

Follow-up discussions are under way to see what restrictions we are facing and how we can overcome them. Access to scientific information and materials exchanges could be greatly facilitated if there were clear international regulations in place.

Without such regulations and support for international networking, progress in biodefence research, especially the expertise needed to develop new prophylactics and tools for rapid detection and containment of diseases, will be significantly hindered.

**Amanda Ozin, Karen Bade,
Stefan H. E. Kaufmann**

Department of Immunology,
Max Planck Institute for Infection Biology,
Schumannstr. 21-22, D-10117 Berlin, Germany

Today is the time to take environmental action

Sir—Your News story highlighting the different perceptions of climate change between US and German audiences of the environmental-disaster film *The Day After Tomorrow* (*Nature* **431**, 4; 2004) raises some interesting issues, and resonates with the findings of our recent survey of UK film-goers. The prime minister, Tony Blair, reaffirmed his commitment in September to the United Kingdom taking the lead in combating climate change. Our findings indicate that there is a clear and urgent need for governments to provide support for individuals who wish to take action.

Our research (see www.tyndall.ac.uk/research/theme3/summary_t3_dat.shtml)

shows that seeing the film did, at least in the short term, change people’s perceptions. Viewers were significantly more concerned not only about climate change, but also about other environmental risks such as biodiversity loss and radioactive waste disposal. However, the portrayal of extreme events in the film also confused people: they believed extreme climate impacts were less likely, and would not be experienced within their lifetime, after seeing the film.

We found that many viewers of the film expressed a strong motivation to act on climate change—more so than before seeing the film. Less than 5% of the 301 people surveyed believed that there was no point in taking action. But despite being strongly motivated, people did not know what action to take. They require specific guidance on what to do to mitigate climate change, with positive images and examples to enable them to make appropriate changes to their everyday lives.

In his speech last month to the Prince of Wales’s Business and the Environment Programme, Blair said: “To make serious headway towards smarter lifestyles, we need to start with clear and consistent policy and messages, championed both by government and by those outside government: telling people what they can do that would make a difference.”

We agree that the time is right to provide incentives to householders, perhaps through domestic transport and energy initiatives, to help translate public support for addressing climate change, as we have seen following *The Day After Tomorrow*, into concrete personal, individual and collective action.

Katrina Brown and the TDAT group

Tyndall Centre for Climate Change Research,
University of East Anglia, Norwich NR4 7TJ, UK

Antibiotics at the crossroads

Are we making the right choices to bring new drugs to the marketplace?

Carl Nathan

The public takes for granted that the pharmaceutical industry can anticipate society's medical needs and meet them. This faith is nowhere more evident than in the expectation that antibiotics are readily available to treat bacterial infections. After all, infectious diseases are the second-leading cause of death worldwide and the third-leading cause of death in economically advanced countries. But surprisingly, despite growing bacterial resistance to existing drugs, antibiotic development in the pharmaceutical industry is steeply declining (see chart, right)¹⁻³. This new problem is converging with an old one — the scarcity of antibiotics to treat diseases prevalent mainly in poorer regions.

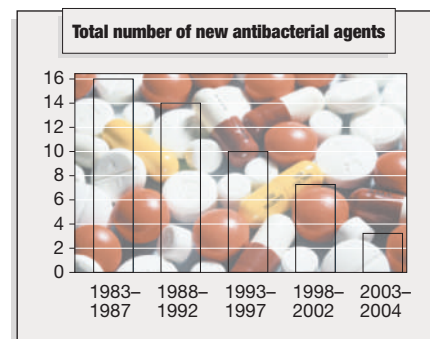
The emerging crisis in wealthy nations and the long-standing crisis in poor nations result from the same causes — economic, regulatory and scientific — each exacerbated by the problem of antibiotic resistance. Government agencies and professional societies have addressed the latter problem³⁻⁶, but little has changed. We need new approaches, beginning with the recognition that the antibiotic crises of wealthy and poor nations are the same. The challenge is this: what can we do about the level of antibiotic research and development, which has long been insufficient to meet the needs of most populations, and now is plummeting? Causes of this decline are reviewed below, followed by 'blue sky' proposals for a more constructive approach to the permanent struggle with infectious disease. The focus is on drugs, but vaccination has a major role to play in reducing dependence on antibiotics.

Economic pressures

With respect to profit margins, financial markets hold the pharmaceutical business to a higher standard than almost any other industry. The demand for blockbuster drugs pressures companies to focus on long-term treatment of chronic conditions in preference to brief treatments for bacterial infections⁷. Most of the antibiotics that major firms make are designed for broad-spectrum activity, so that they can be used by as many patients as possible. This shortens the market life of an antibiotic — as widespread use of an antibiotic hastens the emergence of resistance against it. To ward off resistance, physicians are urged to spare their use. With profits thus restrained in the medical arena, pharmaceutical firms send roughly half their antibiotic output to the



The lack of new drugs leaves children in developing countries especially vulnerable to disease.



food industry⁵. Pork, fowl, fish and dairy producers use antibiotics to maintain stock and foster growth. This selects for resistant bacteria, which can find their way into human populations — hastening the demise of the drug and making once-treatable infections incurable⁴⁻⁹.

Industry's retreat from developing new antibiotics is leading to a loss of expertise in both practical and theoretical aspects of antibiotic biology¹. As industry reassigns or retires its microbiologists, academia will in turn train fewer⁵. When wealthy societies demand a resumption of antibiotic research, it will take years to rebuild the knowledge base.

Better business models

The Global Alliance for Tuberculosis Drug Development, a not-for-profit agency, is building a case to persuade industry that moderate profits can be made by developing

antibiotics for a disease prevalent in poor regions¹⁰. AstraZeneca has opened a research centre for anti-tuberculosis drug development in India (www.astrazenecaindia.com) and GlaxoSmithKline has assembled a team to work on drugs for tuberculosis and malaria (www.gsk.com/financial/rep02/CSR02/GSKcsr-7.htm). Although these are positive steps, these initiatives are not enough to equip us to treat infections endemic in poor regions, nor do they address the emerging shortage of antibiotics for bacterial infections in wealthy nations.

Academic scientists are making rapid advances in the chemical biology of infectious diseases. But they lack access to medicinal chemistry, pharmacology and the expertise to turn 'hits' into drug leads, or those leads into drugs.

What is needed is a new player on the scene: a not-for-profit drug company. The profit sector could provide leadership. Encouraged by tax incentives, industry could give sabbaticals to its scientists and executives to work at a not-for-profit firm in rotation. Many in the pharmaceutical industry would like nothing better than to contribute personally to an endeavor in which their company (as a whole) is constrained from engaging.

A not-for-profit firm could pursue research differently, protecting its intellectual property by filing patents, but also advertising its work openly, with the goal of licensing the intellectual property gratis to any company or agency that commits to produce and distribute the resulting drugs on a basis that would serve the needs of patients and society. For example, distribution in low-income markets could be on a for-cost basis whereas distribution in wealthy markets could remain for-profit.

Biotechnology firms are beginning to 'translate' the ideas of academic researchers into drugs, but it is difficult for small firms to mount a world-class effort at medicinal chemistry and pharmacology, especially now that expertise in antibiotic development is scarce. A not-for-profit drug company could perform these services in exchange for a share of revenues from sales in high-income countries, coupled with a commitment that the drugs be distributed on a not-for-profit basis elsewhere.

The best established way to delay the emergence of antibiotic resistance is to use one or more drugs in combination — known as combination therapy. A potential source of income for a not-for-profit, industry-supported drug company could there-

A. TDR, WHO/SPL

C. V/SPL; SOURCE: CLIN. INFECT. DIS.



AstraZeneca's research operation in Bangalore brings much-needed investment and expertise to India in the search for anti-tuberculosis drugs.

fore be contract work for private firms to identify effective drug combinations at an early stage. In fact, a not-for-profit firm could test the idea that shared knowledge allows early identification of targets (usually microbial enzymes) whose combined inhibition is lethal to the bacteria.

The majority of the funding for a non-profit firm would probably have to come from government and foundations. Tax incentives could encourage the for-profit sector to furnish services in kind or at cost, including equipment, supplies, chemicals, clinical development, and regulatory and legal services. Manufacturing could be contracted to factories in low- and middle-income countries.

One can foresee many problems, such as back-flow of drugs from poor regions to wealthy ones. Such problems will be easier to manage than asking twenty-first century societies to accept nineteenth-century death rates from infection³.

Regulatory obstacles

Another major disincentive in the development of new antibiotics is the current system of regulatory requirements that discriminate against their approval^{1,5}. In the United States, companies must demonstrate that a new antibiotic is superior to existing agents when used against infections caused by drug-sensitive strains. Existing agents are so effective against drug-sensitive strains that a new antibiotic is unlikely to be much better than an older one. Yet testing new antibiotics against infections caused by antibiotic-resistant bacteria is exceptionally difficult, as patients with serious drug-resistant infections have usually been treated with other antibiotics before resistance is documented. In short, the regulatory system is geared to

generic standards of safety and efficacy—it makes no allowance for the specific case of antibiotic resistance. Companies have withdrawn from developing products against which they believe the regulatory system discriminates.

Smarter regulations

In agreement with recent recommendations³ of the Infectious Disease Society of America (see news feature on page 892), I believe regulatory requirements and patent incentives should be revised to encourage the pharmaceutical industry to develop new antibiotics. New antibiotics should be approved if they meet three tests. First, the safety profile is acceptable for the severity of the infection; second, the drug is effective in patients against antibiotic-sensitive bacteria; and third, the drug is effective *in vitro* against bacteria that are resistant to one or more existing antibiotics used to treat that infection. After a new antibiotic is approved, its clinical efficacy should be monitored in patients who are infected with bacteria resistant to previously approved antibiotics. This information should be posted on the Internet as it is collected.

Unless new antibiotics are used in combination, resistance⁸ against them will quickly arise. A new pre-approval test should be developed for antibiotics intended to treat persistent or recurrent infections, such as tuberculosis and malaria, which require sustained administration. The manufacturer should run preclinical tests to determine how the new agent interacts with existing antibiotics used to treat that disease (one from each class). This information, combined with the drug's pharmacokinetic profiles, should help regulators to develop new post-approval requirements for treating patients. First, the manufacturer

should specify lists of agents, at least one of which should be used together with the new drug. Second, the manufacturer should monitor clinical efficacy and incidence of drug resistance when such combinations are used in practice.

Patent life should be extended for antibiotics of new chemical classes directed at new targets², analogous to US patent extensions for drugs developed to treat rare genetic diseases.

In addition, all new antibiotics should be banned from widespread administration to healthy animals. It remains for the rest of the world to embrace an enforceable version of the ban enacted in the European Union in 1998 (ref. 6), or alternatively, to provide tax incentives with the same effect.

Stalled science

It would be simplistic to blame market forces and regulatory requirements alone for the antibiotic crisis. There is another and more surprising cause—industrial research and development has mainly produced variants of older antibiotics, when new drugs are sorely needed. Over the past few decades, only two new chemical entities have entered clinical practice as antibacterial agents, and only one whose target is in a new biochemical class^{1,2,8,9}. It is surprising that the well has gone dry, despite heavy investment to dig it deeper using combinatorial chemistry and computational biology. Although genomic analyses are revealing hundreds of potential targets in pathogens, it remains a fact that almost all agents used to treat bacterial infections either have unknown enzymatic targets or target just four classes of enzymes—those involved in synthesis of protein, nucleic acids, cell walls or folate⁸. How did yield decline while knowledge grew and tools improved?

The answer may lie in a set of premises that were so successful that they hardened into dogmas. In a time of rapid intellectual expansion, dogmas are constraints.

Fresh approaches

There is no longer any reason to confine ourselves to drugs that inhibit the synthesis of protein, nucleic acids, cell walls and folate simply because such drugs have been so successful. We must find new microbial targets. First, synthesis offers too narrow a set of targets. Macromolecules such as DNA and protein have life cycles. Birth need not be the only point of intervention, as processing, repair and degradation are also points of vulnerability. This is the rationale behind efforts to target the proteasome in *Mycobacterium tuberculosis*¹².

Another broad set of targets are the enzymes of core metabolism (intermediary metabolism, energy generation and micronutrient acquisition) in the bacteria. A third set of targets (overlapping the other two) are the enzymes that enable the pathogen to resist the defences of the host. After all, evolution has had more time than scientists to select chemicals to kill pathogens; the host has reactive oxygen intermediates, reactive nitrogen intermediates and pore-forming peptides in its arsenal. Of course, evolution has also strengthened microbial defences against the host's chemistries. But we can aid the host's immunity by using antibiotics that disable pathogens' resistance mechanisms^{12,13}.

The goal of antibiotic development is inhibition of essential enzymes — those the pathogen needs to survive. But survive where? Traditionally, tests to determine what enzymes are essential have been in rich, highly oxygenated culture medium. The conditions facing pathogens in the

host during many infections — especially those that are persistent — can be drastically different from and more demanding than such conditions *in vitro*. Not only do metabolic niches *in vivo* differ from culture broth in many ways (for example, in oxygen, iron, pH and carbon source), but the immune system acts to suppress the pathogen's replication and damage many of its molecules. Successful pathogens adapt by expressing a different set of genes than they do in culture¹⁴. Accordingly, a different repertoire of genes is essential *in vivo* than *in vitro*^{9,11,15}. In short, essentiality is conditional and the



Overuse of antibiotics in livestock has led to an increase in resistant bacterial strains.

conditions defining essentiality are multiple. For example, many infections, including tuberculosis, enter phases of latency — a state of equilibrium between the bacterium and the host response. The agents now used to treat tuberculosis kill rapidly growing bacteria in culture within hours. In contrast, treatment of tuberculosis in people takes at least 6–9 months of daily combination therapy, because the microbial targets that are essential in exponential growth phase may not be so critical during latency or persistence⁵.

Systematic studies in yeast teach us that

mutations in many genes are only lethal when combined with mutations in others¹⁶. This is also true for pathogens. We should therefore target gene products which are essential together, even when they are not essential individually. For example, *M. tuberculosis* has two genes encoding isocitrate lyase enzymes, each of which is capable of supporting lipid

metabolism. Disruption of both genes in combination, but neither alone, causes rapid bacterial decline *in vivo* (J. McKinney, personal communication). Why not target both lyases at once? This idea will require fundamental changes in scientific and regulatory approaches. First, antibiotic development needs to be cooperative, not competitive — an approach that a not-for-profit drug company could pursue using drug candidates from different manufacturers. Second, when a successful combination therapy involves a new unapproved drug, regulatory agencies should allow approval of the combination to

proceed based on clinical tests of the combination itself, rather than insisting on approval for each individual component as they do now.

Another premise that handicaps antibiotic development is that targets in the pathogen must have no equivalent (homologue) in the host. This is to avoid harming the host. Yet most classes of targets inhibited by antibiotics do have host homologues, except those involved in cell wall synthesis. It is time to abandon the premise. Contemporary structural, computational and chemical biology should be able to engineer compounds that can harm the pathogen without harming the host¹⁷.

Conventional antibiotic development has reached an impasse, partly because it demands that new agents have broad-spectrum activity. This imposes severe limitations, as targets must be widely conserved across pathogens — and even then only the most conserved subsites can be targeted. In contrast, it is medically preferable and will preserve the utility of the drugs longer, if antibiotics are highly specific, so that each one is used less often⁸.

Treating infections with pathogen-specific rather than broad-spectrum antibiotics (whenever possible) will require prior, rapid, accurate and specific diagnosis. It makes no sense to use twenty-first century technology to develop drugs targeted at specific infections whose diagnosis is delayed by nineteenth-century methods. Advances in PCR, mass spectroscopy, quantum dot-enhanced immunoassays, nanotechnology, instrumentation and other technologies should be used to develop diagnostics. With further investment, doctors could expect to submit patient specimens (such as throat swabs, blood or urine) to analysis, and receive diagnoses in many cases within minutes to hours. Today, diagnosis usually takes a day or more. Without minimizing the challenge, we should acknowledge that pretreatment diagnosis is key to minimizing the use of broad-spectrum agents and keeping even in the endless race against drug resistance⁸.

It is time to start applying new technologies to antibiotic development. Here are three examples. First, conventional gene disruption in the pathogen does not allow one to test the role of a gene during a given stage of infection, such as latency. If disruption of the gene precludes growth *in vitro*, then the gene-deficient mutant cannot be studied at all⁹. To determine whether a given target is essential *in vivo* we need 'conditional gene inactivation'. This allows the investigator to turn a gene off at a particular time after infection has begun, and thereby model the effect of treating the infection with an antibiotic directed against the gene product.

Second, we must not rely exclusively on screening chemical libraries against enzymes isolated from the bacteria. Although this

The History of Medicine

- 2000 ac — Here, eat this root.
- AD 1000 — That root is heathen. Here, say this prayer.
- 1850 — That prayer is superstition. Here, drink this potion.
- 1920 — That potion is snake oil. Here, swallow this pill.
- 1945 — That pill is ineffective. Here, take this penicillin.
- 1955 — Oops...bugs mutated. Here, take this tetracycline.
- 1960–1999 — 39 more "oops"... Here, take this more powerful antibiotic.
- 2000 — The bugs have won! Here, eat this root.

—Anonymous

Society's ongoing struggle against infectious disease.

approach identifies chemical compounds that inhibit specific targets, it cannot reveal whether they would affect that target inside the bacterium, or even if they would get into the pathogen. We need to identify up front those compounds that can enter the bacterium and inhibit the target in its natural setting. For example, we could replace an endogenous gene encoding a potential target with a 'conditional hypomorphic allele', allowing reduction but not complete elimination of the target (which could kill it). Then we could screen chemical libraries to find compounds to which the 'weakened' mutant is particularly sensitive⁹.

Third, innovative chemistry can allow us to find more potent inhibitors more quickly and cheaply. Drug development usually starts with inhibitors that work at nanomolar concentrations. Conventional screening of compounds rarely yields inhibitors active below the micromolar range. It can take teams of chemists months to make

a micromolar inhibitor a thousand-fold more active. But the target enzyme itself may be able to select weakly binding compounds that are mutually and covalently reactive from two separate but complementary chemical libraries. The enzyme can then catalyse their covalent reaction into a single new compound that inhibits the enzyme with higher affinity¹⁸.

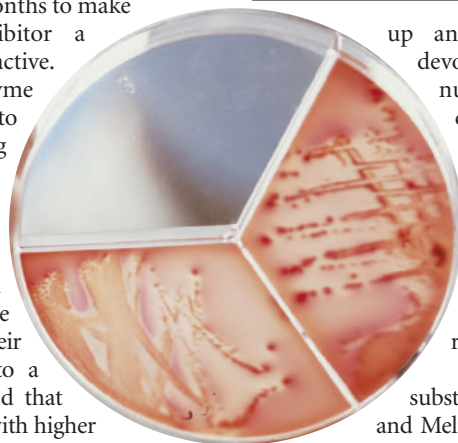
Finally, we must exploit microbial diversity better. Many drug leads have been natural products developed from one bacterial order, Actinomycetales¹⁹. But most of the microbial universe remains unexplored, because most microbial species remain to be cultured. For example, studies of the microbes in soil^{8,19} and water and the viruses that prey on them²⁰ could reveal many compounds and enzymes that may help a given species compete with others in its environment. These natural products can teach us a great deal about microbial vulnerabilities and how to exploit them.

Impossible? Think again

Is it hopelessly unrealistic to imagine not-for-profit drug companies working in a smart regulatory environment, applying fresh scientific approaches to antibiotic development? Perhaps the most challenging aspect of this three-part vision is the notion of a not-for-profit drug company. Yet, at least three models of public-private partnerships for development of anti-infectives²¹ are



New approaches to screening chemical libraries are needed to develop antibiotics.



"It makes no sense to use twenty-first century technology to develop drugs targeted at specific infections, whose diagnosis is delayed by nineteenth-century methods."

up and running. Each is devoted to one or a small number of infectious diseases. Their growing success²² suggests that it is possible to involve private industry in work that society needs, but the market does not competitively reward.

One model, funded in substantial part by the Bill and Melinda Gates Foundation and the Rockefeller Foundation, involves small, not-for-profit drug companies that are virtual and physically distributed. The Medicines for Malaria Venture (www.mmv.org) takes ideas, hits or leads, mainly from academic scientists, and uses a contract mechanism to fund medicinal chemistry and pharmacology in the labs of other academics or pharmaceutical firms²¹. Similar approaches are taken by the Global Alliance for Tuberculosis Drug Development (www.tb Alliance.org) and the Drugs for Neglected Diseases Initiative (www.dndi.org)²¹.

A second model of public-private partnerships involves on-site research and development funded by a major pharmaceutical company in conjunction with a public partner, such as The Novartis Institute for Tropical Diseases. This is funded jointly by Novartis and the Singapore Economic Development Board (www.nitd.novartis.com). A third model, also funded in substantial part by the Gates Foundation, is represented by The Institute for OneWorld Health (www.oneworldhealth.org). This agency uses donated intellectual property to operate

a small, on-site, not-for-profit drug company that prepares vaccines or drugs for malaria, leishmaniasis, trypanosomiasis, helminth infections and diarrhoeal diseases. Finally, the biotechnology industry appears to be positioning itself to contribute to the public-private partnership model through another Gates Foundation-funded initiative, BIO Ventures for Global Health (www.bvgh.org).

All sectors of society, including the pharmaceutical industry, have a major stake in the control of infectious diseases, not only for medical reasons, but also for global economic development and security²³. It is in the interest of both rich and poor societies that initiatives such as those described above grow by orders of magnitude and broaden their scope to include all major infectious diseases that the pharmaceutical industry does not adequately address.

Carl Nathan is in the Department of Microbiology & Immunology, Weill Cornell Medical College, and Programs in Immunology and Molecular Biology, Weill Graduate School of Medical Sciences of Cornell University, 1300 York Avenue, New York 10021, USA.

- Projan, S. J. *Curr. Opin. Microbiol.* **6**, 427–430 (2003).
- Wenzel, R. P. *N. Engl. J. Med.* **351**, 523–526 (2004).
- Infectious Diseases Society of America. *Bad Bugs, No Drugs*; www.idsociety.org/Template.cfm?Section=Home&CONTENTID=7455&TEMPLATE=/ContentManagement/ContentDisplay.cfm (2004).
- Centers for Disease Control and Prevention. *A Public Health Action Plan to Combat Antimicrobial Resistance*; www.cdc.gov/drugresistance/actionplan/html/index.htm (2000).
- Coates, A., Hu, Y., Bax, R. & Page, C. *Nature Rev. Drug Discov.* **1**, 895–910 (2002).
- World Health Organization. *Draft Global Strategy for the Containment of Antimicrobial Resistance*; www.who.int/emc/amr.html (2001).
- Service, R. F. *Science* **303**, 1796–1799 (2004).
- Walsh, C. *Nature Rev. Microbiol.* **1**, 65–70 (2003).
- Miesel, L., Greene, J. & Black, T. A. *Nature Rev. Genet.* **4**, 442–456 (2003).
- Global Alliance for Tuberculosis Drug Development. *The Economics of TB Drug Development*; www.tb Alliance.org/7_6D_Publications.asp (2001).
- Chopra, I., Hodgson, J., Metcalf, B. & Poste, G. *J. Am. Med. Assoc.* **275**, 401–403 (1996).
- Darwin, K. H., Ehrst, S., Gutierrez-Ramos, J. C., Weich, N. & Nathan, C. F. *Science* **302**, 1963–1966 (2003).
- Bryk, R., Lima, C. D., Erdjument-Bromage, H., Tempst, P. & Nathan, C. *Science* **295**, 1073–1077 (2002).
- Schnappinger, D. et al. *J. Exp. Med.* **198**, 693–704 (2003).
- Sassetti, C. M., Boyd, D. H. & Rubin, E. J. *Mol. Microbiol.* **48**, 77–84 (2003).
- Tong, A. H. et al. *Science* **303**, 808–813 (2004).
- Fidock, D. A., Rosenthal, P. J., Croft, S. L., Brun, R. & Nwaka, S. *Nature Rev. Drug Discov.* **3**, 509–520 (2004).
- Manetsch, R. et al. *J. Amer. Chem. Soc.* **126**, 12809–12819 (2004).
- Keller, M. & Zengler, K. *Nature Rev. Microbiol.* **2**, 141–150 (2004).
- Liu, J. et al. *Nature Biotechnol.* **22**, 185–191 (2004).
- Nwaka, S. & Ridley, R. G. *Nature Rev. Drug Discov.* **2**, 919–928 (2003).
- Vennerstrom, J. L. et al. *Nature* **430**, 900–904 (2004).
- World Health Organization, Commission on Macroeconomics and Health. *Macroeconomics and Health: Investing in Health for Economic Development*; www.cmhealth.org (2001).

Acknowledgements

Thanks to A. Apt, K. Deitsch, H. Djaballah, B. Ganem, W. Jorgensen, T. Kapoor, M. MacCoss, V. Mizrahi, S. Projan, L. Quadri, K. Rhee, M. Rosenberg, D. Russell, D. Scheinberg, D. Schnappinger, D. Tan, T. Templeton and P. van Helden for stimulating discussions. Special thanks are due to P. Davies, S. Ehrst, M. Glickman, B. Kelly, J. McKinney and S. Nwaka.

The evolutionist's tale

A journey back through our evolution with a growing band of ancestors.



ILLUSTRATIONS BY CHRISTIAN DARKIN

The Ancestor's Tale: A Pilgrimage to the Dawn of Life

by Richard Dawkins

Weidenfeld & Nicholson (UK)/Houghton Mifflin (US): 2004. 528 pp. £25/\$28

Jerry A. Coyne

Richard Dawkins' success as a popularizer of science lies in his remarkable ability to convey complex, subtle and technical issues without dumbing them down or patronizing the reader. It also helps that he writes beautifully and, as is evident from the titles of his books (think of *The Selfish Gene* or *The Blind Watchmaker*), has an exceptional gift for using metaphor to capture and solidify abstract ideas. Until now, Dawkins has dealt with concepts — most notably natural selection — for it is in the explication of ideas that he excels. In *The Ancestor's Tale*, however, he leaves his familiar literary niche. Instead of ideas, he traffics here in facts: the history of life on Earth. Facts, however, don't need metaphors. How does Dawkins handle this new direction in his work?

Perhaps conscious that his arrival in 'fact-land' must differ from that of previous travellers (two other Richards — Fortey and Southwood — have recently published splendid overviews of life's history for the general reader), Dawkins uses a pair of gimmicks to enliven his parade of information.

First, he writes his history in reverse. Beginning with modern humans and moving backwards in time, he describes our lineage as we successively join — a geneticist would say coalesce — with the common ancestors of other species. Human evolution has involved 40 such joints, each occupied by what Dawkins calls a "concestor", and each is the subject of a single chapter. He begins, of course, with our common ancestor with chimps, followed by the concestor with

gorillas, then other primates, and so on through the fusion with early mammals, sponges, plants, Eubacteria and ultimately the *Ur*-species, probably a naked molecule of RNA. This narrative is engagingly written and attractively illustrated with reconstructions of the concestors, colourful phylogenies, and photographs of bizarre living species. The book is also remarkably up to date and, despite its size, nearly error-free. Especially notable are Dawkins' treatments of human evolution and the origin of life, the best accounts of these topics I've seen in a crowded literature.

Second, lacking more familiar metaphors, Dawkins adopts a literary conceit, modelling the book on Geoffrey Chaucer's *The Canterbury Tales*. This parallel itself takes two forms. Dawkins views species as pilgrims marching into the past, joining each other genetically on a 3-billion-year journey to evolution's Canterbury: the first "replicator". In addition, he fleshes out his phylogeny with 58 sidebars: chaucerian "pilgrim's tales" told by extant species, each describing a biological concept or method.

How well does this literary parallel work? The answer is mixed. The reverse-history tactic is useful, emphasizing as it does our common ancestry with other groups, but it is somewhat anthropocentric, following only one thread — ours — in the tapestry of life. The pilgrims' tales, on the other hand, represent a return to terra firma for Dawkins, as they deal largely with ideas. These stories are clearly the centrepiece of the book, not only for the reader, but seemingly for Dawkins himself. (One can almost sense his relief as he interrupts his dutiful account of amphibian evolution to explain the evolutionary significance of "ring species" in salamanders and seagulls.)

Although Dawkins aficionados will rec-

ognize some tales from his earlier books, most are new and absorbing. The grasshopper's tale, for example, uses geographic variation within orthopteran species to inspire a remarkably sensible treatment of human racial variation, a topic whose political overtones usually drive biologists to panicked circumlocution. The peacock introduces us to sexual selection, the flounder to evolutionary imperfection, and the rotifer to the disadvantage of reproducing sexually. A lot of arcane but engrossing biology lies buried in these tales: we learn, for example, why some electric fish swim without undulation (such movement would distort the electric fields that aid navigation and prey detection), and why anteaters are the only animals that do not digest their food with hydrochloric acid (formic acid from ants does the job nicely).

Occasionally, however, the combination of history and pilgrims' tales fails to cohere, and both elements sometimes have a whiff of the textbook. Several digressions, such as the exhaustive explanations of log-log plots, radiometric dating and molecular clocks, are simply too long and complicated. These stories are familiar to scientific readers, but are likely to fatigue rather than inspire the laity. It is not that this material is dull — Dawkins could make a field guide lively reading — but that it sometimes lacks the intellectual spark that fired his previous works.

In view of the book's breadth and the author's strong opinions on many issues, any biologist can find something to criticize. My main gripe is Dawkins' strong emphasis — apparently influenced by the work of Geoffrey Miller — on sexual selection as a likely engine for the evolution of important human traits, including a large brain and bipedal gait. Dawkins suggests, for instance, that ancestral females favoured males who were smarter and adopted an attractive

vertical posture. But unless such traits improve the fitness of both males and females (in which case natural selection alone can do the job), sexual selection will produce differences between the sexes, as it probably did with human body size. Why, then, don't we see knuckle-walking women with chimpanzee-sized brains? Although we may never know why humans became erect and brainy, sexual selection seems among the least plausible of many alternative theories.

Despite its considerable merits, *The Ancestor's Tale* seems the least compelling of

Dawkins' works. Given the author's talents, however, this may be akin to judging *Coriolanus* Shakespeare's least compelling play. Thankfully, Dawkins returns to top form in the final chapter, a philosophical overview of the extraordinary events he's just recounted. Here he grapples with questions of whether evolution would produce similar creatures if it began anew (a qualified yes), and whether evolution itself promotes "evolvability", making species even more likely to respond to future selection (another qualified yes).

His final sentences are quintessential

Dawkins, with language at its most lambent and elegant, used simply to express profound truths: "The fact that life evolved out of nearly nothing... is a fact so staggering that I would be mad to attempt words to do it justice. And even that is not the end of the matter. Not only did evolution happen: it eventually led to beings capable of comprehending the process, and even of comprehending the process by which they comprehend it." ■

Jerry A. Coyne is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

First steps in science

Curious Minds: How a Child Becomes a Scientist

edited by John Brockman

Pantheon: 2004. 256 pp. \$23.95, £16.99

Paul L. Harris

John Brockman asked 27 distinguished scientists from diverse fields, including physics, psychology, biology, mathematics and robotics, to write a short biographical sketch, describing their childhood and what got them hooked on science. The heterogeneity of the essays they wrote serves as an antidote to easy speculation about there being any essential precursors to scientific accomplishment.

Is a scientifically oriented family important? It looks that way at first as you read through the opening piece by psychologist Nick Humphrey. One of his early memories — vividly described — is of being taken by his grandfather on Boxing Day to dissect a frog in the empty, silent anatomy department of University College, London. The grandfather, as it happens, has a Nobel prize in physiology, and is just one family figure in a distinguished panoply. Humphrey candidly writes: "What I gained from this childhood environment was a sense of intellectual entitlement — a right to ask questions, to pry, to provoke, to go where I pleased in pursuit of knowledge."

That sense of intellectual entitlement is echoed by cognitive scientist Alison Gopnik. She writes of her parents' devotion to their children's intellectual lives, not with a view to "enrichment" or "achievement", but with a gift for making such intellectual activity "the accepted, ordinary, happy way that civilized people went about their lives". Ray Kurzweil, an inventor, recalls a similarly intense intellectual life in his family circle: "The intense and animated discussions were invariably about new ideas, usually those of intellectuals I had never heard of. The way for me to get attention was to have an idea."

But consider, by contrast, the childhood



of the computer scientist Jaron Lanier. His mother died when he was nine and he was raised by an indigent father with an interest in psychic phenomena. For a while they lived in tents in southern New Mexico before moving into a fragile geodesic dome designed by his father. Such an eccentric upbringing might nurture intellectual independence, but it's a long way from the scientific dynasty described by Humphrey or the high-minded but nurturing family in which Gopnik was raised. Neuroscientist Joseph LeDoux describes an equally unpropitious start. His father was a bull rider in the rodeo and hoped that his only child would emulate him.

Another plausible candidate for later scientific accomplishment is a keen childhood

interest in some branch of the natural world: insects, stars, fossils or electricity. Some contributors do display such a passion. Computer scientist Rodney Brooks, for example, was already fascinated by electric circuits at the age of seven. And Lynn Margulis, a biologist, writes of lying on her belly to watch ant colonies. Yet there are disconcerting exceptions. Another biologist, Richard Dawkins, tells an amusing story about mistaking a blue tit for a chaffinch. Shocked at the boy's error, his grandfather turned to Dawkins' father and remonstrated: "Good God, John — is that possible?"

The conclusion that does emerge from this collection is that a huge number of different paths lead to a successful scientific career. Some can be tentatively traced back

to early childhood, but others receive a bump start later in life from a mentor or teacher, and many illustrate how chance encounters provoke a dramatic new direction. Without providing any general conclusions, these essays offer the idiosyncrasy and curiosity value that we expect of good, narrative history, combined with much fine writing.

Two contributors — Stephen Pinker and Judith Harris — adopt a scientific stance toward the whole narrative enterprise. Don't trust our childhood recollections, they caution, and certainly beware of any causal connections that we infer. Psychological research on the unreliability of childhood memories and our frequent lack of introspective access to potent influences on our choices lend weight to their cautionary remarks. But few readers will feel comfortable reading these accounts so sceptically. Many of the contributors revisit their childhood with gusto — and with palpable curiosity. What they write is quirky, absorbing and persuasive in just the way that good stories are. ■

Paul L. Harris is at the Harvard Graduate School of Education, 503A Larsen Hall, Appian Way, Cambridge, Massachusetts 02138, USA.

Stalking life's second secret

The Proteus Effect: Stem Cells and their Promise for Medicine

by Ann B. Parson

Joseph Henry Press: 2004. 304 pp. \$24.95

Lee M. Silver

On 28 February 1953, Francis Crick famously burst into a Cambridge pub with Jim Watson in tow and exclaimed triumphantly: "We have found the secret of life!" The rest, as the cliché goes, is history. Both scholarly and popular accounts of the personalities and events surrounding the discovery of the double helix, and subsequent advances in molecular genetics, have appeared in tens of thousands of books and articles. The DNA story is taught in school biology classes to eight-year-old children. Several DNA-based Nobel prizes have already been won, and more are no doubt waiting in the wings.

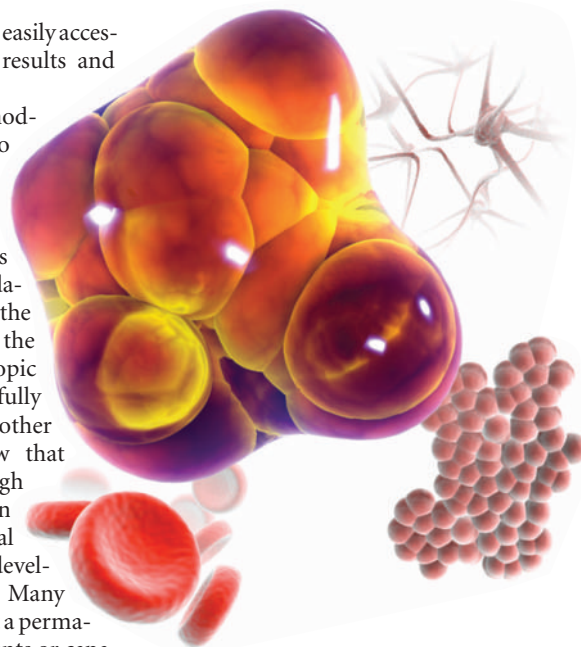
Genes and genomes, however, represent only one of two pillars of likely twenty-first-century progress in biomedicine. The second is stem-cell biology. And yet, while everyone knows of Watson and Crick, hardly anyone has heard of Leroy Stevens and other early stem-cell pioneers. Indeed, many scientists are unaware of the five decades of advances that occurred before the 1998 isolation of human embryonic stem cells. In her new book *The Proteus Effect*, journalist Ann

Parson fills the gap with a breezy, easily accessible narrative of the people, results and ideas that have shaped the field.

Parson traces the birth of modern stem-cell research back to 1953, the same year that the double helix made its debut. Although the means for composing and reproducing life's genetic recipes had been a fundamental secret of life, it was not the only one. A second mystery was the process by which a microscopic embryo could develop into a fully functional human being or other mammal. Embryologists knew that development occurred through the controlled growth, division and specialization of individual cells, but they didn't know how developmental control was asserted. Many assumed, without evidence, that a permanent loss of biological components or capabilities was responsible for the narrowing of cellular potential. By implication, it seemed that a postembryonic mammalian cell should never be able to reverse course and return to an embryonic state. The exceptions were the germ cells (eggs and sperm), which still had the potential to retrace normal development only after combining with each other.

Goal-oriented unidirectional development was easily accepted as scientific dogma because it aligns neatly with traditional Judaeo-Christian-derived religious doctrine, which permeates Western culture. As recently as 2002, Princeton professor Robert George — a defender of strict Catholicism and member of President Bush's Council on Bioethics — claimed the mantle of science when he wrote that the direction of a human embryo's growth "*is not extrinsically determined*, but is in accord with the genetic information *within* it ... It is clear that from the zygote stage forward, the major development of this organism is *controlled and directed from within*, that is, by the organism itself" (George's emphases). Consequently, according to George and others on the council and in the Bush administration, human zygotes and blastocysts (newly fertilized embryos that have just started to divide) are "already living human beings" deserving of protection from the murderous pursuits of biomedical scientists. Perhaps George would come to realize how outdated his views of embryology are if he read Ann Parson's book (and then again, perhaps not).

In the winter of 1953, Leroy Stevens was searching for morphological differences among inbred mice at the Jackson Laboratory in Bar Harbor, Maine, when he chanced upon the observation of an enlarged testicle in about 1% of adult males of the 129 strain. The cause of the enlargements was an inde-



pendently growing mass of a type previously found only in human patients. It was called a teratoma because of its monster-like characteristics. Human teratomas are typically covered with skin and hair, and contain a bizarre diversity of tissues and organs normally found in other parts of the adult body, including synaptically connected neurons, muscle, beating heart tissue, bone, teeth and sometimes eye-like entities or whole limbs. Stevens realized that the 129 strain's predisposition to develop teratomas gave him a unique tool for tackling the mystery of their origin, which could provide insight into normal developmental processes.

Two years and 17,000 animals later, Stevens had dissected his way back through younger and younger mice, and ultimately fetuses, to discover that random genital-ridge cells occasionally flew off-course and developed like the innards of early embryos before exploding into teratomas. In other strains, females were predisposed to teratoma formation. Ovarian cells morphed, by themselves, into an organism that Stevens said "looked just like a normal embryo, but then it would all get mixed up. It became disorganized and you had a tumour instead of a baby mouse." Stevens didn't think that mutant genes were the main trigger of this aberrant development, and when he placed normally fertilized embryos into unusual environments, such as the kidney capsule, they also became teratomas instead of mice.

Teratomas are not ordinary tumours. Along with their adult tissues, some contain cells that can grow and divide for years without becoming specialized. When dispersed and injected into mouse abdomens, these cells developed into "thousands of small structures resembling 5- or 6-day mouse embryos floating freely". Stevens was convinced that such "embryonic carcinoma"

cells had the capacity to settle down and participate in completely normal tissue and organ development if given a chance in a normal embryonic environment. The embryologists Ralph Brinster, Beatrice Mintz and Richard Gardner proved him right by producing chimaeric mice derived from a mixture of normal embryo cells and embryonic carcinoma cells.

If this were the whole story, teratomas would have remained no more than interesting medical oddities. But, in 1981, Martin Evans and Matthew Kaufman of Cambridge University, and Gail Martin at the University of California, San Francisco, independently

grabbed the reins of development when they extracted embryonic stem cells directly from mouse embryos and kept them growing in culture indefinitely without differentiating. In 1998, the same process was perfected for use with human embryos, and the age of regenerative medicine was initiated. The goal now is to discover the factors and conditions that will transform embryonic-like cells into whatever tissue is required to overcome a particular disease or human condition.

Parson engages the debate between supporters and opponents of human embryo research by allowing the main players to speak for themselves. She doesn't advocate

for or against, although the book's subtitle leaves no doubt as to her own position. The final chapter provides a balanced assessment of the therapeutic potential of stem cells in both the short and the long term, disease by disease and organ by organ. All in all, Parson admirably brings to life the stem-cell story from a tiny Maine fishing village to the battle for the American presidency in 2004. ■

Lee M. Silver is a professor of molecular biology and public affairs at Princeton University, Princeton, New Jersey 08544, USA. He is the author of the forthcoming book Biotechnology in a Spiritual World (Ecco Press).

Food for thought for geneticists

Why Some Like It Hot: Food, Genes and Cultural Diversity

by Gary Paul Nabhan

Island Press; 2004. 223 pp. \$24

T. Colin Campbell

It's not that we are what we eat, proposes Gary Nabhan, but rather, we are what foods our ancestors ate and what genes they gave to us. In *Why Some Like it Hot*, Nabhan argues that natural selection and other evolutionary processes mediated by food choices unique to each geoclimatic and cultural domain have played important roles in generating human genetic diversity.

According to Nabhan, the time period for this evolution is likely to have been much shorter than is generally supposed. Our unique gene profiles evolved within the past few thousand years or so, perhaps as mutations caused by secondary compounds in staple foods and culinary herbs that were indigenous to certain geoclimatic regions. These genomic profiles could have been moulded by food choices, and by pressure from endemic diseases, to reduce the risk of disease.

This view departs from the darwinian idea that our current gene pool resulted from a very slow process of random selection to maximize individuals' survival. The fact that 99.9% of our genome is shared by all humans, present and past, is cited as evidence of an exceptional genetic stability that arose many millennia ago, perhaps during and before Palaeolithic times.

This observation of early genetic similarity and stability has suggested to some observers that there is a one-size-fits-all diet, perhaps one that was commonly used during Palaeolithic times, but Nabhan challenges this view. He points out, as others have, that of the 3 billion nucleotides in the human genome, 3 million of them (0.1%) have remained in play and can be used to create genetic diversity. A change of only one

nucleotide can make a substantial difference.

Nabhan identifies a group of 26 'disease genes' that are likely to have been fashioned by food factors and endemic diseases. He cites adult-onset diabetes, lactose intolerance and heritable food allergies as examples of interactions between genes, food and disease. He goes on to say that a large number of us are subject to one or more of these genetic 'disorders', as some would call them — indeed they are so common that they should be considered normal.

He discusses in considerable depth the extensively studied link between malaria, sickle-cell anaemia, the consumption of fava beans, and glucose 6-phosphate dehydrogenase (G6PD) deficiency, in order to illustrate how a careful study of biology, culture and history can be much more rewarding than

one of these disciplines alone. He takes the reader on a trail of discovery, visiting the Mediterranean island of Sardinia, where malaria has long been endemic. Here the traditional springtime consumption of fava beans offers those with the genetic disorder of G6PD deficiency some protection against the mosquito-borne disease. Coupling cultural, biological and historical analyses in this way is the basis for the field of nutritional ecogenetics.

Nabhan also cites examples that tend to be found in regions of the world, particularly islands, where the resulting interactions became reasonably well established and remained stable. For example, on the island of Crete and elsewhere in the Mediterranean region, a high-fat diet is associated with a lower risk of heart disease than would be



expected from studies of other Western subjects. His travels to these regions and his discoveries are presented as engaging personal experiences. He ends his book with a particularly notable example of the remarkable clinical experiences of Terry Shintani and colleagues, who have studied the food sensitivities now suffered by indigenous Hawaiians as they adopt a modern Western diet.

The story told by Nabhan is thought-provoking. He implies that each of us, as individuals, should consider our food choices and their health effects with reference to our own evolutionary past. However reasonable this argument seems, it leaves unanswered the critical question of how such information can be used to create contemporary dietary advice for the public. Most people now are a heterogeneous mixture of genes and food habits that will be virtually impossible to untangle in a way that can tell us what we should be eating to be healthy. Even if we knew the evolutionary pedigree of our contemporary genes — which is most improbable — we would be hard put to match our genetic predispositions with the specific foods likely to make us most healthy. I dread to think what the marketplace will make of this account.

In decrying the one-size-fits-all diet proposal promoted by some dietary advocates, Nabhan seems to ignore the remarkable nutritional range and food choices that exist within such supposedly monolithic diets. For example, the recommended low-fat, whole-food, plant-based diet neither connotes an unvarying diet nor implies exactly the same health benefits for all consumers. It allows for, and even encourages, the consumption of a wide variety of such foods. But the diversity for different regions and different groups of people can still be used to generate most of the same health benefits. This is the beauty of the work of Shintani and his colleagues, who have produced remarkable health benefits when obese and diabetic native Hawaiians are re-introduced to their native whole-plant-based foods, which are low in fat and high in fibre and antioxidants. In a similar way, a range of different plant-based foods can be used to control or even reverse a variety of serious diseases.

It is remarkable that meaningful genetic adaptation can occur in a few thousand years or so, but even more so that genetic-like adaptations can occur within only one or two lifetimes. Dietary experiences before or shortly after birth, either direct or maternal, can imprint substantial biochemical and morphological changes that become stabilized well into adulthood as if they were genetic — yet they are more likely to be post-transcriptional or post-translational. Such is the nature of dietary adaptation, a process that is continually at work, short term and long term, minimizing harm and

making the most of what food and other resources are available.

Despite these minor criticisms, the book is well worth reading, for it should stimulate an important debate about what constitutes dietary adaptations and sensitivities. ■

T. Colin Campbell is in the Division of Nutritional Sciences, 307 Savage Hall, Cornell University, Ithaca, New York 14853-6301, USA.

The sincerest form of flattery

Imitation of Life: How Biology is Inspiring Computing

by Nancy Forbes

MIT Press: 2004. 176 pp. \$25.95, £16.95

John Doyle and Marie Csete

Generations of engineers have recognized that, in many respects, biology does it better. *Imitation of Life* is a whirlwind history, richer even than its subtitle suggests, through various computational disciplines inspired by biology. This is an ambitious undertaking for such a short book but, although it ignores some important unifying principles, its brevity is also a virtue. The inspirations from biology are scattered throughout the book, and their collective impact is felt best when the book is digested whole, at one sitting. The early chapters on biology as a metaphor are the least satisfying, and any reader who stops there may never return for the genuine delights that follow.

Some historians argue that the inspiration provided by avian feathers and flapping may have delayed heavier-than-air flight for centuries. Critics would argue that artificial neural nets, genetic algorithms, cellular automata and artificial life, which are covered in the first four chapters, are the modern

equivalents of flapping, whereas advocates see them as mature fields no longer in need of review. Nancy Forbes is generous about their successes, but does little to resolve the issue. Only later are there brief discussions of hierarchy, modularity, layers of control, and system architecture — key concepts in computing that could help to inform a more thorough analysis.

In contrast to the earlier sections, which use biology as a metaphor, the chapters on DNA computing and biomolecular self-assembly describe the direct use of biological chemistry or materials to create technological artefacts that have little or nothing to do with biology. Forbes is clearly more interested in these topics, and this enthusiasm may well spread to readers; the section on the intriguing computational power in the organization of DNA is particularly well presented. These chapters start to make it clear that understanding biological principles in some depth is an essential part of profitable imitation. Making DNA computing work requires a firm grasp of the principles and careful design, but now anyone can download cellular automata or genetic algorithm software and run laptop artificial-life experiments.

The chapters on amorphous computing, computer immune systems and biologically inspired hardware further underscore the idea that, as biology is better understood, inspiration can proceed more from mimicry than metaphor, and contribute more directly to solving difficult computational tasks. What makes this work (and these chapters) more compelling is the fact that engineers in these fields have crossed disciplinary lines to gather a deep and practical understanding of biology and biological experimentation. The recent explosion in our detailed knowledge of biology, and the glimpses this provides of its organizing principles, has considerably enriched biologically inspired computing.

The final chapter reverses direction and



looks at biology through the lens of computer science and electrical engineering. This is conceptually the deepest chapter but its brevity limits it to a few well-chosen examples. The technologies of any age have always provided metaphors for biology — from myths about our origin involving dust and clay to the industrial revolution's hydraulic and, later, electromagnetic imagery, from telegraphs to telephones and finally to computers. The book contains almost no explicit discussion of complexity, and this omission is particularly noticeable here. Much of computer science is about organizing complexity, from 'very large-scale integrated' circuit design to object-oriented programming and the layered protocols of the Internet.

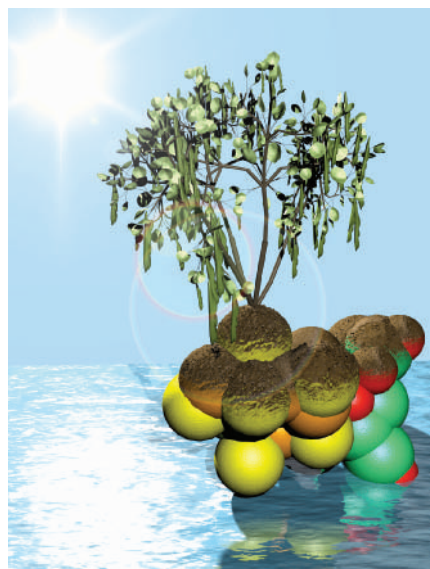
The history of flight is again instructive. By the nineteenth century, engineers had realized that lift, drag and propulsion were the key fundamental mechanisms, and toy gliders became commonplace. Yet only with the Wright brothers' insight that active control was needed for steering and compensation for uncertainties did flying literally take off. By the 1940s, unpiloted aircraft had demonstrated fully automatic transatlantic flight and landing, and some engineers now argue that flight would generally be safer without pilots.

Similarly, the vast majority of computers now are 'embedded', with automated sensing, control and actuation, all entirely hidden during normal operation. Computer control systems such as these represent both the main use of computers and the main source of complexity in technological systems, but are barely mentioned either in this book or elsewhere. This is a pity, as they are the points of greatest contact between engineering and biology. Biologists would have benefited from a discussion of sensing and adaptation in computation and networking, such as Internet routing and congestion control, because sensing and adaptation are widespread in biology, for example in the immune system. Perhaps biologically inspired computing is not yet at a 'Wright brothers' stage, with many fundamental mechanisms just emerging, old superficial metaphors being set aside, and systems-level integrating concepts remaining murky. Hopefully, the Wright brothers of biologically inspired computing are among the many fascinating characters described by Forbes, and their subjects will take off as promised.

This book is easily accessible but is probably most suited to, and beneficial for, biologists, as a clearly written, non-technical primer describing activities on the other side of campus. Biologists will have to ignore some unfortunate simple errors (such as neurons called axons, AIDS as an autoimmune disease and vaccines made from weakened antigens) but can easily read around them. Nonetheless, this text helps to bridge a daunting technical language barrier

and should facilitate further dialogue between biologists and computer scientists. ■

John Doyle is at the California Institute of Technology, Pasadena, California 91125, USA, and Marie Csete is at Emory University School of Medicine, Atlanta, Georgia 30322, USA.



The roots of nitrogen fixation

The World's Greatest Fix: A History of Nitrogen in Agriculture

by G. J. Leigh

Oxford University Press: 2004. 254 pp.

£20, \$29.95

Vaclav Smil

The discovery by Fritz Haber of a method for fixing ammonia from its elements led to the development of the modern nitrogen-fertilizer industry. The growing applications of nitrogen compounds in agriculture have had enormous demographic, economic and environmental consequences. Most histories of nitrogen fixation focus on this part of the story, but in *The World's Greatest Fix*, G. J. Leigh provides a more evenly distributed account. Nearly two-thirds of Leigh's text is devoted to general considerations of nitrogen's chemistry and agronomy, and of technical developments before Haber's discovery.

The book begins with a brief introduction to nitrogen fixation and its importance for agriculture, before tracing the development of agronomic practices in pre-Colombian America (by the Aztecs and Mayas), dynastic China and the Roman Empire, with particular attention to classic agricultural accounts (and the value of manure) by Cato the Censor, Columella, Pliny the Elder and Varro. Next, the story advances to the modern era, focusing on English farming from Roman times to the beginning of scientific

agriculture in the seventeenth century, before moving on to the trade in guano (bird droppings are a rich source of nitrogen) and Chilean nitrates (originally Bolivian and Peruvian) — England was the leading importer of these sources of nitrogen. Leigh then takes us into the laboratory, dealing with the alchemy of nitre and the early chemistry of nitrogen (from Paracelsus to Lavoisier and Chaptal), the birth of agricultural chemistry (thanks to Davy, von Liebig and Boussingault) and the discovery (by Hellriegel and Wilfarth) that microorganisms and some plants can fix their own nitrogen.

Leigh then describes the evolution of the first commercial methods invented to fix nitrogen. The Norwegian arc process, which combines N_2 and O_2 in an electric arc furnace and uses the resulting nitric oxide to produce HNO_3 , was made possible by inexpensive hydroelectricity. The synthesis of cyanamide, by reacting CaC_2 with N_2 , also fixed nitrogen but was energy-intensive. Next come Haber's experiments, beginning in 1903, and the quest, led by Carl Bosch, to commercialize them, beginning at the BASF plant in Ludwigshafen, Germany, in 1909. Here I found the only factual errors worth noting: the German chemist who formulated the effective catalysts needed to run the ammonia synthesis was Alwin (not Alois) Mittasch. This work also led to the synthesis of methyl alcohol and the hydrogenation of coal. And Mittasch did not collaborate with Haber in 1903 (at the time he was Ostwald's assistant in Leipzig); he joined BASF in 1904, working under Bosch.

The penultimate chapter explores the continuing mystery of biological fixation, the author's main area of research interest — he spent most of his professional life at the Unit of Nitrogen Fixation, which was set up in the mid-1960s at the University of Sussex. He describes the discovery of nitrogenases — the enzymes that fix nitrogen in both free-living and symbiotic bacteria — and their modes of action; he discusses the structures made up of iron, molybdenum and sulphur that are at the heart of these remarkable molecules; and he reviews the unsolved puzzles regarding their active sites.

In closing, Leigh reviews the scale and effects of nitrogen-fertilizer use, focusing on aquatic eutrophication — in which excessive nitrogen promotes algal growth and the subsequent depletion of oxygen — and nitrates and human health. The health effects of nitrates are frequently exaggerated, but eutrophication, which Leigh treats rather lightly, still does not get enough attention, given its severity and increasing occurrence. There is also now considerable eutrophication of terrestrial ecosystems.

I always like unusual tit-bits, asides and images, and Leigh's book has its share of them. Among my favourites is a reference to Humphrey Davy's discussion of various

fertilizers, including pilchards in Cornwall. There is also a photograph of Tyntesfield, an impressive Victorian Gothic Revival mansion built near Bristol between 1863 and 1866, financed largely with money from the guano trade. And there is a description of a frustrated British inspection of the world's first ammonia plant in Oppau, Germany, in 1919 with the intention of reverse-engineering it in Britain. Leigh conveys a great deal of information in 220 pages of text, and does so in an easy-to-read, clear and accurate style. This is altogether a fine book.

Those who would like to know more about the key protagonist of the nitrogen story, yet are unable to read German, can finally get an (abridged) English translation of Dietrich Stoltzenberg's massive 1994 biography *Fritz Haber: Chemist, Nobel Laureate, German, Jew* (Chemical Heritage Foundation, 2004). ■

Vaclav Smil is in the Department of Environment, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.



Freedom from smallpox?

The Life and Death of Smallpox

by Ian Glynn & Jenifer Glynn

Profile Books (UK)/Cambridge University Press (US): 2004. 288pp. £17.99/\$25

Hugh Pennington

"The world and all its people" have "won freedom from smallpox", declared resolution WHA 33.3, signed on 8 May 1980 at the eighth plenary meeting of the 33rd World Health Assembly. The assembly was right to congratulate itself on the success of a campaign by its parent body, the World Health Organization, to eradicate smallpox as a "natural" disease.

There had been many difficulties. Money was often short. There were sceptical experts whose influence had to be countered. In the final stages of the programme, the virus went to ground in Somalia and Afghanistan, not the easiest places to hunt it down. In many ways, the virology was the easiest part — despite its reputation, smallpox couldn't spread as quickly as other viruses, such as measles, so it could be stopped in its tracks by isolating victims and using vaccination to create cordons sanitaires. It was also easy to track because of its highly visible and distinctive rash. Ali Maow Maalin, a hospital cook in Marka, Somalia, was the last person to receive a natural case of smallpox, developing his rash on 17 October 1977. The campaign had finally been successful.

But the people of the world have not "won freedom from smallpox". Unlike the dodo and the dinosaur, it is not extinct: it survives

in frozen form in laboratories at Atlanta, Georgia, and Koltsovo in the Novosibirsk region of Russia. And it also survives as a threat in the minds of bioterrorism experts and the politicians they advise, as there is a worry, unproved, that Koltsovo might have been leaky — they consider that the value of smallpox to a terrorist lies not in its lethality but in its ability to panic the public.

The Life and Death of Smallpox explains why. It starts by describing the recorded impacts of the disease in ancient times, and how it replaced plague as an epidemic killer. The book's focus then shifts to the immunological attack on the virus, with the appearance of Lady Mary Wortley Montagu. She was an early promoter in England of variolation — the deliberate infection of the young with smallpox from a mild case — having seen it in operation in Turkey.

Jenifer Glynn is a historian, so although the account of Lady Mary is somewhat anglocentric, it contains nice touches. Lady Mary lost her eyebrows after an attack of smallpox that damaged her beauty and, no doubt, stimulated her interest in the disease. And we learn of her elopement with Edward Wortley Montagu to avoid an arranged marriage with Clotworthy Skeffington. But the overall account of variolation is certainly not anglocentric. As the book records, variolation persisted right up to the end of natural smallpox, changing from a practice that had some merit in the eighteenth century (its mortality was lower than that of randomly acquired disease) to one that helped the wild virus to persist in the mid-1970s.

Edward Jenner is the hero of the book, of course. Most of the chapters are devoted to describing his introduction of vaccination,

its spread and its successes, finishing with the eradication campaign, which depended on it. More generally, the trials and tribulations of vaccination also feature in the book. Britain is currently having trouble with its immunization policy, for example. Because of a perception that the combined measles-mumps-rubella vaccine may be linked to autism, some parents are rejecting it, leaving their children unprotected. Others are opting for unofficial alternatives. *The Life and Death of Smallpox* provides an excellent account of similar events a century and more ago. A group of people opposed to vaccinations was active then and, in 1898, compulsory vaccination in England was abandoned — 'conscientious objectors' to vaccination could apply for an exemption certificate.

As a historical account of smallpox, the Glynn's book is excellent. It attends to vaccination controversies and personality clashes in a masterly fashion. Virus-related policy events in the aftermath of the terrorist attack of 11 September 2001 are covered well. But those who look for a definitive account of the origins of vaccinia, the vaccine virus, will be disappointed. To virologists this will be no surprise, because its history is shrouded in mystery and its molecular genealogy is puzzling. As a former poxvirologist, I was a little disappointed not to read more about the molecular and cell biology of the virus itself, but its virological quirks and peculiarities, and those of its relatives, are striking and complex enough for another book. This one is for the general reader. ■

Hugh Pennington is at the University of Aberdeen Medical School, Foresterhill, Aberdeen AB25 2ZD, UK. He is president of the Society for General Microbiology.

Climbing the co-evolution ladder

Co-evolution: Earth history involves tightly entwined transitions of information and the environment, but where is this process heading?

T. M. Lenton, H. J. Schellnhuber
and E. Szathmáry

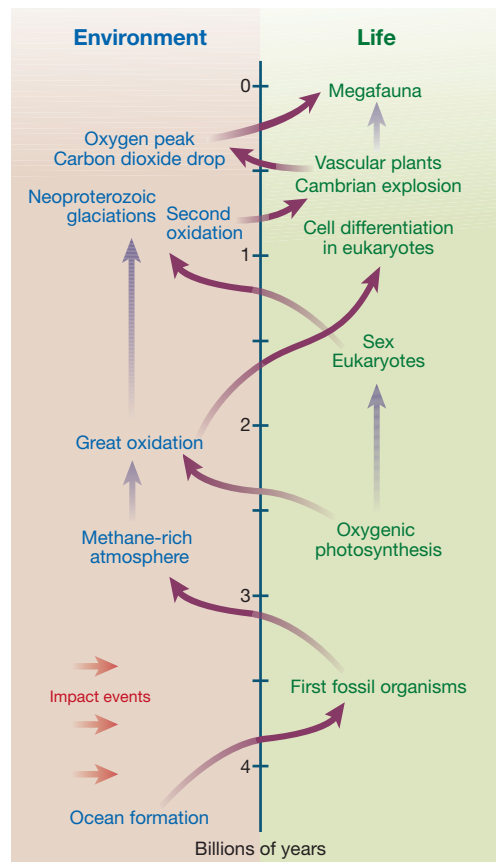
Stanislav Lem's science-fiction masterpiece, *Solaris*, tells the gripping — and scary — story of a super-intelligent super-organism that has transmuted into a vast ocean covering most of the surface of a distant planet. Thus information-processing (that is, active) life and force-driven (that is, passive) environment have finally merged into a single entity.

Here on Earth, we have yet to reach this vanishing point of evolutionary history. But modern civilization already perturbs — if not dominates — various large-scale processes and components of the planet. Most notably, the global 'metabolism' (the cycling of essential elements, including carbon, nitrogen, phosphorus and sulphur) and features of the global 'anatomy' (the landscape textures of the habitable continents) are largely a product of relentless socio-economic action.

Within the framework of Earth system analysis, this can be perceived as the latest step on the grand co-evolutionary ladder of entwined transitions of information and environment. Global industrialization, particularly since the Second World War, induced the transition into the Anthropocene. Its environmental consequences may in turn provoke a transition to an even higher form of worldwide socio-political organization.

Extending our concept of entwined transitions further back in time, we soon encounter the rise of the 'hydraulic societies' in the valleys of the Nile, Euphrat, Tigris and Indus, which were probably founded in response to the great drying of the African–Asian regions that took place in the sixth millennium before present (BP). Before this, organized *Homo sapiens* hunting caused mass extinctions of the prehistoric fauna. Language provided the novel inheritance system that allowed cumulative cultural and technological evolution, and a society resting on complex, negotiated division of labour. Further back, the evolution of hominins in the East African Rift valley was shaped by a rapidly changing environment.

Entwined environment–information transitions have characterized Earth's evolutionary history since its beginning some 4 billion years ago (4 Gyr). Life emerged remarkably soon after surface conditions became habitable, with the formation of oceans and cessation of sterilizing asteroid impacts. The first organisms would have



Earth's co-evolutionary ladder, as built so far...

drained the environment of energetically and structurally useful compounds and replaced them with degraded waste products, including methane. An ultimately dull fate for life, eking out a meagre existence on a lifeboat Earth, was averted when closed recycling loops developed, in which one life form's waste became another's food. These loops are large-scale manifestations of the auto-catalytic nature of the cell, locked in as the core of the global 'metabolism' that is still with us.

Despite recycling, life remained energetically limited until the origin of oxygenic photosynthesis, sometime before 2.7 Gyr. This breakthrough in metabolic evolution greatly increased the free energy supply to the biota, giving life a truly global environmental impact. It facilitated the great oxidation of the atmosphere around 2.2 Gyr, but — as the long time lag indicates — other factors were required. Perhaps oxidation had to await tectonically driven changes in Earth's 'anatomy', including the appearance of shelf seas where reduced organic carbon could reach the sediments and be buried.

Although the energetic stage was now set for global dominance by eukaryotes, the emergence of a soft cell-boundary membrane coupled to an internal skeleton and a means for cellular division were also required. These transitions are thought to have been especially difficult, as they required the fixation of thousands of rare mutations.

Eukaryotes may be implicated in the worst crisis of past co-evolution: the extreme Neoproterozoic glaciations of 0.8–0.6 Gyr that were accompanied by a second rise in oxygen. Whenever eukaryotes started to colonize the land surface, there would have been strong selection for traits that accelerated weathering to access rock-bound nutrients. Weathering of silicates would have inadvertently drawn down atmospheric carbon dioxide and cooled the planet, and weathering of phosphorus would have increased global productivity and contributed to oxygen rise. The latter opened the door for the diversification of larger, hard-shelled, animal life in the Cambrian explosion. After that, the triumph of vascular land plants, causing a further rise in oxygen and fall in carbon dioxide, played its part in creating the environmental conditions in which active megafauna (including ourselves) evolved.

Pursing this concept of entwined evolution may reveal where we are ultimately heading — towards Solaris, or something even scarier.

T. M. Lenton and H. J. Schellnhuber are in the Tyndall Centre, UK, and the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. E. Szathmáry is at the Collegium Budapest, 2 Szentharomsag utca, H-1014 Budapest, Hungary.

FURTHER READING

Schellnhuber, H. J. *et al.* (eds.) *Earth System Analysis for Sustainability* (Dahlem Workshop Reports 91, MIT Press, 2004).

Lovelock, J. E. *The Ages of Gaia — A biography of our living Earth* (Oxford Univ. Press, 1988).

Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Freeman, 1995).

Lenton, T. M. & Watson, A. J. *Geophysical Research Letters* **31**, L05202 (2004).

Schellnhuber, H. J. *Nature* **402**, C19–C23, (1999).

Crutzen, P. J. *Nature* **415**, 23 (2002).

Based on the 91st Dahlem Workshop on Earth System Analysis for Sustainability, Berlin, 25–30 May 2003, with thanks to our fellow participants.

End of the beginning

Lincoln D. Stein

Just over three years ago, it was announced that a first draft of the human genome sequence had been completed. Gaps and errors remained, but the job of fixing those problems is now largely done.

This issue of *Nature* features an article¹ entitled “Finishing the euchromatic sequence of the human genome”. It has been authored by members of the International Human Genome Sequencing Consortium (IHGSC), and appears on page 931. The article marks the latest, but by no means the last, milestone in this historic project. But readers can be forgiven for being a bit confused by the announcement. Wasn't the human genome ‘finished’ several years ago?

The answer is ‘yes’ — and ‘no’. Early in 2001, the duelling IHGSC (public) and Celera Corporation (private) groups published papers in *Nature*² and *Science*³ describing the completion of so-called ‘draft’ sequences. These sequences have revolutionized molecular biology by largely eliminating the need to clone and sequence genes involved in human health and disease. Instead of going to the bench, biologists now go to the web to look up gene sequences in public online databases.

But despite their immediate usefulness, the draft sequences were far from perfect. Both drafts were missing some 10% of the so-called ‘euchromatin’ — the gene-rich portion of the genome — and some 30% of the genome as a whole (which includes the gene-poor regions of ‘heterochromatin’). The drafts contained hundreds of thousands of gaps, and had misassembled regions where portions of the genome were flipped or misplaced. As a result, any large-scale analyses of the genome, such as studies of the mechanisms of gene evolution or the long-range structure of the genome, had to contend with numerous uncertainties and artefacts. For example, studies of ‘pseudogenes’, the dying remnants of genes that have accumulated mutations that render them non-functional, had to contend with the possibility that any apparent pseudogene was instead the result of a sequencing error.

Since the publication of the drafts, the IHGSC sequencing centres have quietly undertaken a laborious ‘finishing’ process, in which each gap in the draft was individually examined and subjected to a battery of steps involving cloning and resequencing stretches



of DNA. The sequence announced today has just 341 gaps remaining, and consists of contiguous runs of sequence averaging 38 million base pairs. The authors estimate that the finished sequence covers 99% of the euchromatic portion of the genome and that the overall error rate is less than 1 error per 100,000 base pairs. This substantially exceeds the original goals for the project.

The finishing procedure roughly doubled the total time and cost of the project. Does it contribute anything new to our understanding of the genome? It does indeed, and to prove the point the authors of the current paper¹ describe several large-scale analyses of the genome that would have been difficult to perform on the draft sequence. One analysis studied the processes of gene birth and death. The authors find 1,183 human genes that show evidence of having been recently ‘born’ by a process of gene duplication and divergence. They also find 37 genes that seem to have recently ‘died’ by acquiring a

mutation that rendered the gene non-functional. The resulting pseudogene then slowly degrades and disappears.

In a second analysis, the authors use the finished sequence to map out segmental duplications — large regions of the genome that have duplicated in recent evolution. They find that 5% of the genome is involved in segmental duplications, and that the distribution of these regions varies widely across the chromosomes. Knowing the nature and extent of such duplications is important for understanding the evolution of the human genome, and for studying the many medically relevant disorders that are involved in segmental duplications, such as DiGeorge syndrome and Charcot-Marie-Tooth syndrome.

Another paper in this issue, by She *et al.*⁴ (page 927), directly compares the outcomes of this second analysis with results obtained on an unfinished version of the human genome (an improved version of the Celera draft). She *et al.* find that the draft version artefactually ‘simplifies’ the genome by eliminating many duplicated regions. Their results bear on one of the highly publicized differences between the public and private

genome projects. The public project used an older strategy in which the genome was first cloned into bacterial artificial chromosomes (BACs); the clones were then mapped, and each clone was sequenced and their sequences assembled individually. Celera championed an untested technique, ‘whole-genome shotgun’ (WGS), in which the entire genome was shattered into bite-size pieces, sequenced, and then assembled by software in one conceptually simple step.

Celera proved that the WGS technique is both technically feasible and provides a dramatic cost-saving over the clone-by-clone approach. Although the Celera draft has languished because the availability of public data in free online databases undermined the company’s business plan to sell genome-database subscriptions, the effort left a permanent mark on the public project. Almost all genome-sequencing projects since then have used some form of WGS. The cautionary results contained in the new

TY IMAGES

papers from the IHGSC¹ and She *et al.*⁴ argue for a hybrid strategy in which WGS is supplemented by a modest amount of BAC cloning and mapping. This would protect draft WGS sequences from some of the 'simplification' reported by She *et al.* and provide the clones needed for finishing selected regions of special interest.

What is next for the human genome project? Even with a finished sequence in hand there is much still to do. Surprisingly, one task is to develop the definitive catalogue of protein-coding genes. In the current paper¹, the number is estimated to be between 20,000 and 25,000. This wide range reflects limitations to state-of-the-art gene-prediction software that leave doubts about the validity of many predicted genes. One promising approach is to use comparative genomics to align the human genome with the genomes of other animals. Because natural selection ensures that functional regions are more highly conserved than non-functional ones, this approach highlights candidate protein-coding regions. The same approach shows promise for finding other functional elements such as gene promoters, which control the timing and level of expression of genes, and micro-RNAs, which have been implicated as regulatory agents of many developmental processes.

Much farther in the future is the task of sequencing the remaining 20% of the genome that lies within heterochromatin, the gene-poor, highly repetitive sequence that is implicated in the processes of chromosome replication and maintenance. The repetitiveness of heterochromatin means that it cannot be tackled using current sequencing methods, and new technologies will have to be developed to attack it. So don't be shocked to see another paper announcing the 'finishing' of the human genome in 2010 — it will describe how the heterochromatin problem has been cracked.

In sequencing the human genome, researchers have already climbed mountains and travelled a long and winding road. But we are only at the end of the beginning: ahead lies another mountain range that we will need to map out and explore as we seek to understand how all the parts revealed by the genome sequence work together to make life.

Lincoln D. Stein is at Cold Spring Harbor

Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

e-mail: lstein@cshl.edu

1. International Human Genome Sequencing Consortium *Nature* **431**, 931–945 (2004).
2. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
3. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
4. She, X. *et al.* *Nature* **431**, 927–930 (2004).

split into two distinct evolutionary lineages: the actinopterygians (ray-finned fish), which include teleosts such as pufferfish and zebrafish, and the sarcopterygians (lobe-fins), which include lungfish, coelacanths and ourselves (Fig. 1). By matching up the genes on each pufferfish chromosome to the related genes on human chromosomes, Jaillon *et al.* deduce that the extinct ancestor had 12 pairs of chromosomes. Work on partially completed genome sequences had suggested this number^{4,5}, but the new analyses add fascinating detail to the picture. For example, it is now possible to say which genes were on which chromosomes, despite this unknown animal having been extinct for more than 400 million years.

One puzzling observation concerns the apparent stability of the genomes of ray-finned fish. It seems that the ancestral genome underwent as few as ten large inter-chromosomal rearrangements (exchanges, fissions or fusions) to give rise to the present-day *Tetraodon* genome. Indeed, 11 *Tetraodon* chromosomes have not experienced any such rearrangements. Only one human chromosome (14) can make the same claim, despite the timescale being identical.

Although the genomes of ray-finned fish may have been slowly evolving in terms of chromosome breakages and fusions, they have experienced a cataclysmic event in their history. Jaillon and colleagues' analyses of the complete *Tetraodon* genome sequence show clearly that a duplication of the whole genome occurred sometime within the ray-finned-fish lineage. This inference is not new, having been previously suggested from analyses of the Hox-gene clusters and other gene families in zebrafish, *Takifugu* and other teleosts^{4–7}, but the conclusion has remained controversial⁸.

Two new analyses should now settle the issue, however. First, Jaillon and colleagues plotted the chromosome positions for about 750 pairs of 'ancient' duplicated genes within the *Tetraodon* genome, revealing related pairs of chromosomes or chromosomal regions. Every chromosome is involved, consistent with an ancient whole-genome duplication. In the second test, chromosome positions for more than 6,000 pufferfish genes were compared with the positions of related genes in the human genome. This revealed a striking pattern of 'double conserved synteny', meaning that one chromosomal region in humans matches two in pufferfish, across the entire genome. This is a clear echo of whole-genome duplication in the ray-finned-fish lineage. Every gene, on every chromosome, was duplicated, although there has since been a massive degree of gene loss and local gene shuffling.

When did this whole-genome duplication occur? Analysis of zebrafish genetic maps strongly suggests that this species also underwent such an event in its history⁴.

Comparative genomics

Small genome, big insights

John Mulley and Peter Holland

The genome of a second pufferfish species has been sequenced. Why is this important? Because comparing this genome with that of other animals yields a wealth of information on genome evolution.

It is still early days for the field of comparative genomics. Only around a dozen species of animal have so far had their complete DNA sequence determined, even to draft coverage. These are predominantly the widely studied model species, such as mice, fruitflies and nematode worms, or species of particular interest to humans, such as the malaria-carrying mosquito.

It may come as a surprise, therefore, to find that the list now includes not one, but two species of Tetraodontiformes, a relatively obscure group of fish also known as puffers. Following on from the publication two years ago of the genome sequence of the Japanese pufferfish *Takifugu rubripes*¹, Jaillon and colleagues² report, on page 946 of this issue, the near-complete sequence of the spotted green pufferfish *Tetraodon nigroviridis*. *Takifugu* is a poisonous marine fish best known to connoisseurs of sushi restaurants, whereas *Tetraodon* is a small, brackish-water pufferfish commonly kept in aquaria. But, like all Tetraodontiformes, the

two species share a feature of great convenience for genomics: their cells possess less DNA than those of any other group of back-boned animals — about eight or nine times less than human cells.

Although the *Tetraodon* genome is small compared with that of other vertebrates, sequencing it was still a hugely formidable task. The research reported in this issue² was performed in a collaboration between Genoscope in France and the Broad Institute of the Massachusetts Institute of Technology and Harvard University in the United States. Together they have generated a genome sequence of impressive accuracy and coverage, with 64% of the DNA sequence mapped to specific chromosomes³.

By comparing the *Tetraodon* genome sequence with that of humans, Jaillon *et al.* even allow us to peer into the genome of the last common ancestor of pufferfish and humans — a primitive bony fish that lived hundreds of millions of years ago. The descendants of this long-extinct ancestor

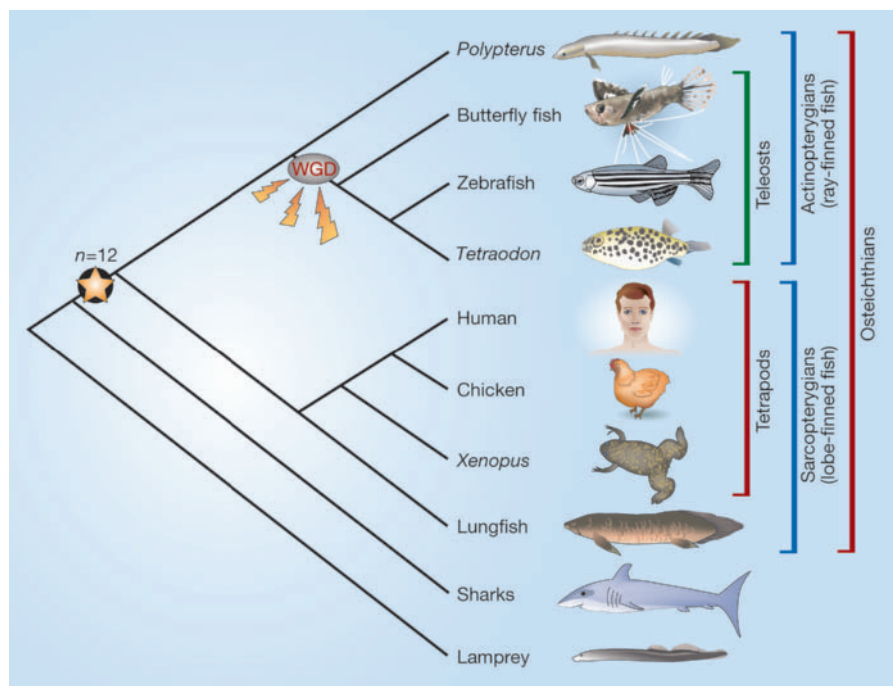


Figure 1 Evolutionary relationships between humans, pufferfish and other vertebrates. Jaillon *et al.*² have sequenced the genome of the pufferfish *Tetraodon nigroviridis*. By comparing this genome sequence with that of humans, the authors deduce that the extinct ancestor of actinopterygians (ray-finned fish, including pufferfish) and sarcopterygians (lobe-finned fish, the lineage that gave rise to humans) had 12 pairs of chromosomes ($n=12$). They also show that a whole-genome duplication (WGD) occurred during the evolution of ray-finned fish.

Pufferfish and zebrafish belong to distinct taxonomic orders of fish, so the duplication must have occurred early in teleost evolution. As previously pointed out⁷, this implies that traces of the ancient whole-genome duplication should be found in more than 20,000 species of living teleost fish. But teleosts do not make up the whole of the ray-finned fish. Significantly, a study of one of the Hox-gene clusters of an earlier (more 'basally') branching ray-finned fish, *Polypterus* (Fig. 1), found no evidence of a genome duplication⁹. Together with data from other basal actinopterygians¹⁰, this suggests that the genome duplication occurred close to the origin of the teleost fish themselves, perhaps 230 million years ago.

Less clear are the biological consequences. It is tempting to suggest that the species richness of the teleosts is somehow related to the whole-genome duplication, either because natural selection has 'exploited' the extra genes, or because differential mutation of duplicate genes caused reproductive isolation, facilitating speciation¹¹. However, much of teleost diversity is found in just one group, the acanthopterygians ('spiny fins'), which underwent a massive increase in diversity only around 55 million years ago. So if the whole-genome duplication did affect species richness, it was not immediate, and further studies of morphological and genetic evolution in teleosts will be needed to resolve the mechanisms involved.

A final lesson from the *Tetraodon* study² concerns the power of comparative genomics, both for gaining insights into mechanisms of genome evolution and for deducing genome organization in extinct species. But we have a long lineage of extinct ancestors, which means that a wide range of genomes will need to be compared if we want to look at each node in our evolutionary history (Fig. 1). Particularly useful will be complete genome sequences for a shark, a lamprey and amphioxus, as each will provide insight into yet more ancient ancestral states. We may not have long to wait: this year the Joint Genome Institute in California began sequencing the amphioxus genome, while the National Human Genome Research Institute has announced plans to sequence that of the sea lamprey.

John Mulley and Peter Holland are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. e-mail: peter.holland@zoology.oxford.ac.uk

1. Aparicio, S. *et al.* *Science* **297**, 1301–1310 (2002).
2. Jaillon, O. *et al.* *Nature* **431**, 946–957 (2004).
3. www.genoscope.cns.fr/externe/tetranew
4. Postlethwait, J. *et al.* *Genome Res.* **10**, 1890–1902 (2000).
5. Naruse, K. *et al.* *Genome Res.* **14**, 820–828 (2004).
6. Amores, A. *et al.* *Science* **282**, 1711–1714 (1998).
7. Taylor, J., Braasch, L., Frickley, T., Meyer, A. & Van de Peer, Y. *Genome Res.* **13**, 382–390 (2003).
8. Robinson-Rechavi, M., Maranchi, O., Escriva, H. & Laudet, V. *Curr. Biol.* **11**, R458–R459 (2001).
9. Chiu, C.-H. *et al.* *Genome Res.* **14**, 11–17 (2004).
10. Hoegg, S., Brinkmann, H., Taylor, J. & Meyer, A. *J. Mol. Evol.* **59**, 190–203 (2004).
11. Lynch, M. & Force, A. G. *Am. Nat.* **156**, 590–605 (2000).



100 YEARS AGO

The Cultivation of Man. The author of this book is very much in earnest. He condemns modern civilisation in strong terms for its many vices, especially for its worship of money and the mammonite marriages that result from it, and urges that men should apply to their own species the methods of the breeder of cattle. He recommends polygamy, apparently in all seriousness, and not as a mere counsel of perfection. It would, of course, destroy the family, but to this Mr. Witchell has no objection... Certainly he speaks out fearlessly, and that is no small merit. But it is to be regretted that he did not study his subject more before writing. "Natural selection," he says, "is sometimes operative, chiefly among the poor." Considering that in England nearly fifty per cent. of the population die before the average age of marriage, this is a wonderful understatement. If we bear the facts in mind, we can hardly agree with Mr. Witchell that the business man is "the surviving type," *i.e.* apparently the type that is to survive to the exclusion of others. Business men are not a separate species. There is a continual upward movement of able men from the great underlying social stratum, and from this stratum directly or indirectly our successful men, as we call them, have emerged.

From *Nature* 20 October 1904.

50 YEARS AGO

Anatomist, pathologist, epidemiologist, sanitarian and clinician, and one of the most advanced thinkers in the history of the medical sciences, Giovanni Maria Lancisi was born in Rome three hundred years ago, on October 26, 1654... It was at [Pope] Clement's request that in 1707 he wrote his monumental treatise "De subitaneis mortibus", in which he carefully records the pathological lesions of the brain and heart observed at autopsy, gives the first description of syphilis of the heart and of growths on the valves, and lists hypertrophy and dilatation of the heart as a cause of sudden death. Lancisi's book, "De motu cordis et aneurysmatibus" (1728), is another landmark in the history of heart disease, for it stresses the significance of heredity, syphilis and violent emotions as causes of aneurysm... In the tercentennial year of his birth he is gratefully remembered chiefly for having laid the foundation for a true understanding of the pathology of the heart.

From *Nature* 23 October 1954.

General relativity

Frame-dragging confirmed

Neil Ashby

According to a prediction of general relativity, the spinning mass of the Earth affects the motion of satellites. A measurement of this 'frame-dragging' effect confirms Einstein's theory.

General relativity predicts that a spinning mass distorts space-time — one of a variety of 'gravitomagnetic' phenomena that are absent in Newtonian gravity. Unfortunately, gravitational forces are so weak that it is useless to try to detect this warping of space-time unless the mass is very large and spinning rapidly — an astronomical body, say, such as the Earth or the Sun, or a neutron star. After many years' work, and through the analysis of millions of 'laser-ranging' measurements made from more than 50 Earth-based stations, Ciufolini and Pavlis¹ have confirmed the twisting effect of the spinning Earth on the orbits of two artificial satellites (page 958 of this issue). The remarkable precision they have achieved is due in large part to recent improvements in the modelling of Earth's gravitational field.

According to general relativity, a spinning flywheel imparts a twist to space and time in its proximity that can affect a nearby

gyroscope. If a frictionless gyroscope is placed near the flywheel's axis of rotation, the gyroscope's spin axis will be dragged along in the direction of the flywheel's rotation. However, if the gyroscope is placed near the flywheel's perimeter, its spin axis will be dragged in a direction opposite to the flywheel's rotation. First analysed by Joseph Lense and Hans Thirring, soon after Einstein published his general theory of relativity, the phenomenon is known as the Lense–Thirring effect — or 'frame-dragging', because the spin axes of ideal gyroscopes could be used to track the directions of coordinate axes in an inertial reference frame.

Frame-dragging is one aspect of the class of relativistic phenomena loosely known as gravitomagnetism, through their analogy with the effects of ordinary magnetic forces on moving electric charges. General relativity describes gravity in terms of a set of ten independent quantities (the components of

a second-rank tensor). The gravitomagnetic terms, however, vanish unless the mass producing the gravitational field is moving. Among the effects caused by gravitomagnetic forces are precessions (similar to the wobble of the axis of a spinning top) of the orbital plane of a satellite around a spinning body (such as Earth), or precessions of the orbit of an Earth satellite as the Earth–satellite system orbits the Sun. This latter effect is usually called de Sitter or geodetic precession.

Precise measurement of these effects predicted by relativistic gravity theories is crucial, as they have important implications for our view of the cosmos. Gravitomagnetic effects can be significant in many astrophysical systems. For example, in the binary pulsar B1913+16, geodetic precession of the orbits may cause the pulsed beam from this star to precess out of our line of sight in a few dozen years, and to reappear some centuries later². However, the uncertainties in such distant systems are usually too large for astronomical objects to be used for precision tests of gravitomagnetic effects.

In the case of the Earth–Moon system, the Moon's orbit may be thought of as a gyroscope with its spin axis perpendicular to the Moon's orbital plane. A change in the direction of this axis can be measured by observing the change in the nodal angle; this is the angle between an astronomical reference line (the first point of Aries, in Earth's equatorial plane) and the line of intersection between Earth's equatorial plane and the Moon's orbital plane. According to general relativity, the geodetic precession of the nodal angle is only a little more than five-millionths of a degree per century, which amounts to motion of the node (on the orbit) of about 3 metres per month. This prediction has been verified³ to an accuracy of about 0.7%, through many years of accurate laser ranging — bouncing laser signals off retroreflectors placed on the lunar surface by Apollo astronauts — and by the development of extremely sophisticated models of the perturbed orbital motion and rotations of the Moon. However, frame-dragging of the Moon's orbit caused by the spin of the Earth is negligible, because the effect falls off rapidly with distance and the Earth is not spinning particularly fast.

But near-Earth satellites suffer larger perturbations owing to gravitomagnetic effects. These perturbations are still extremely tiny, and require enormous effort to measure, as Ciufolini and Pavlis¹ have shown. The two satellites used in their analysis are LAGEOS and LAGEOS 2 (for Laser Geodynamics Satellite). LAGEOS was launched in 1976; the satellite is compact, completely passive, and consists of a heavy sphere covered with retroreflectors. LAGEOS 2 was launched in

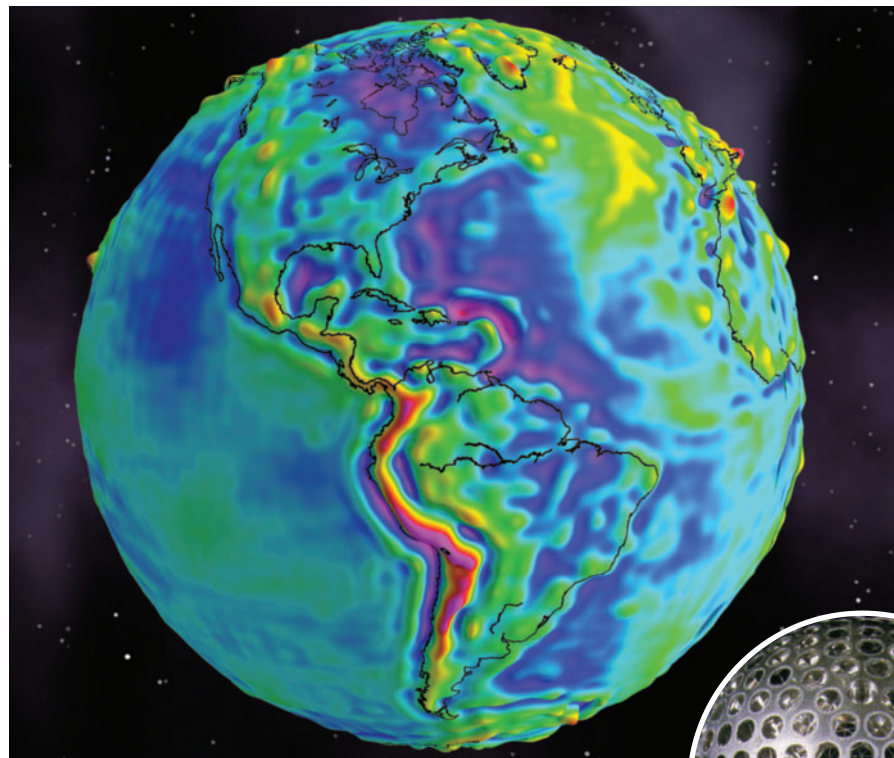


Figure 1 General relativity and Earth's gravity field. The gravity field is not spherical, as the relief on this map measured using the GRACE satellites shows. These features must also be factored into tests of 'frame-dragging', an effect predicted by the general theory of relativity. Ciufolini and Pavlis¹ have used measurements of the changing orbit of the LAGEOS satellites (one shown in inset) to confirm frame-dragging, or, more formally, the Lense–Thirring effect. Their measurement, to an accuracy of 10%, is the most precise so far.

1992 from the Space Shuttle. Accurate laser ranging to these satellites is used to develop better models of Earth's gravity field. This field is not spherical, but has many ripples because of the variations in Earth's mass distribution (Fig. 1). As a result, any frame-dragging experiment must take account of large orbital perturbations that are a solely due to ordinary Newtonian gravity.

In fact, uncertainties in some of the ripples in Earth's gravity field are large enough to mask frame-dragging effects almost completely: for a satellite at an altitude of 12,000 km, frame-dragging of the node is only about 1.9 metres per year, whereas the nodal precession due to oblateness of the Earth is many thousands of kilometres per year. Geodesy missions such as CHAMP and GRACE^{4,5}, whose purpose is to further refine models of Earth's gravity field and study crustal motions, have been essential in reducing the uncertainties. Ciufolini and Pavlis¹ have used two observable parameters — the nodal precession rates of both LAGEOS satellites — to eliminate the principal uncertainty caused by Earth's oblateness. From the remaining data, which consist of more than 100 million laser-ranging observations between Earth and the satellites, the frame-dragging effect could be measured.

The result — the first reasonably accurate measurement of frame-dragging — confirms the general theory of relativity to within a conservative error estimate of 10%. Further analysis is anticipated as additional geodesy missions are undertaken to improve our knowledge of Earth's gravity field. And results are eagerly awaited from NASA's Gravity Probe B, which was launched into polar orbit in April this year. This satellite had been under development for more than 40 years, and seems to be performing well. Its array of gyroscopes and readout systems should produce an even more precise measurement of gravitomagnetic precession, to an accuracy better than 1%. Precise observations of such theoretical predictions will improve our understanding of the cosmos. In all experiments performed up to now, general relativity has been accurately confirmed.

Neil Ashby is in the Department of Physics, University of Colorado, 2000 Colorado Boulevard, Boulder, Colorado 80309-0390, USA.

e-mail: ashby@boulder.nist.gov

1. Ciufolini, I. & Pavlis, E. C. *Nature* **431**, 958–960 (2004).
2. Kramer, M. *Astrophys. J.* **509**, 856–860 (1998).
3. Williams, J. G., Newhall, X. X. & Dickey, J. O. *Phys. Rev. D* **53**, 6730–6739 (1996).
4. <http://earthobservatory.nasa.gov/Library/GRACE/grace2.html>
5. Reigber, C. et al. *J. Geodyn.* (in the press).

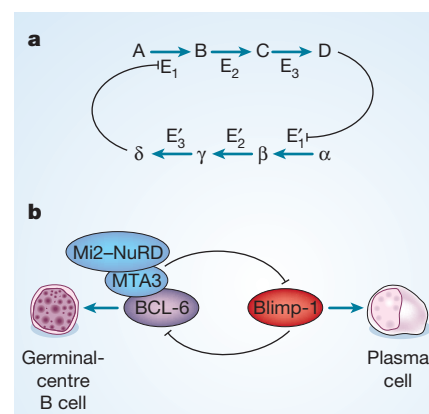


Figure 1 Double-negative circuits in biology. a, In metabolism. Two enzyme pathways are shown. The end-product of one pathway, D, inhibits the other pathway by inhibiting enzyme E'_1 . Conversely, the end-product of the second pathway, δ , inhibits the first pathway by inhibiting E_1 . Transient changes in the concentrations of D or δ , perhaps because of changes in environmental conditions, can swing this circuit towards increased production of one end-product or the other. Adapted from ref. 1. b, In B-cell differentiation. BCL-6 and Blimp-1 are two gene repressors that inhibit each other's expression. As Fujita *et al.*² now show, the MTA3–Mi2–NuRD complex contributes to the repressive activity of BCL-6. BCL-6 is expressed in germinal-centre B cells, Blimp-1 in plasma cells. Transient changes in the expression or activity of either repressor can push a cell towards one or the other differentiation state.

Cancer

Negative feedback for B cells

Louis M. Staudt

The discovery of a protein that regulates the production of antibody-generating B cells has implications for our understanding of how cancers of the immune system develop — and how they might be treated.

Some 40 years ago, Jacques Monod and François Jacob¹ described the role of double-negative regulatory circuits in biological systems. In their example, two enzyme pathways each yield an end-product that inhibits the activity of the other pathway (Fig. 1a). This circuit has the useful property that a perturbation of one of the regulators, even transiently, can push the system towards one of two cellular states. Although the model was developed by studying microbes, Monod and Jacob envisaged that such circuits would be ideally suited to the differentiation (specialization) of animal cells. Lacking the appropriate experimental tools to test this idea, they noted with some exasperation that “eventually, however, differentiation will have to be studied in differentiated cells”. Now, writing in *Cell*, Fujita and colleagues² describe a new component of an elegant double-negative circuit that regulates the differentiation of B cells into antibody-secreting plasma cells. Excitingly, their findings might bear on the study and treatment of cancer.

B cells, also known as B lymphocytes, are a crucial weapon in our immune armoury. When these cells meet a particular foreign molecule (antigen), they enter a microenvironment known as the germinal centre in lymphoid organs. There, they divide prolifically and diversify their antigen receptors by a specialized ‘hypermutation’ mechanism. The process produces some B cells whose receptors have a particularly high affinity for their cognate antigen; these cells are selected for differentiation into plasma cells — the cellular factories that secrete antigen-specific protective antibodies.

A double-negative circuit involving the gene-repressor proteins BCL-6 and Blimp-1 has previously been postulated to regulate this differentiation step^{3,4}. BCL-6 is found in germinal-centre B cells, binds to the *Blimp-1* gene, and inhibits its activity⁵. Conversely, the Blimp-1 protein is found in plasma cells, and directly or indirectly represses the *BCL-6* gene, along with virtually every other gene that is expressed only in germinal-centre B cells³ (Fig. 1b). Blimp-1 also represses

the cell-division gene *c-myc*, leading to the proliferation arrest that is characteristic of plasma cells⁶. Inhibiting BCL-6 in cultured germinal-centre B cells ‘de-represses’ Blimp-1, resulting in differentiation into plasma cells⁴. In germinal centres, strong stimulation of a B cell by its antigen or by activated T cells may transiently decrease BCL-6 expression^{7,8}, thereby tipping the balance in favour of Blimp-1 expression and differentiation into plasma cells.

Fujita *et al.*² now describe an important new component of this regulatory network — the MTA3 protein, a cell-type-specific subunit of a complex called Mi2–NuRD. This complex helps to repress gene expression by influencing the way in which DNA is ‘packed’ in chromosomes. Somewhat serendipitously, Fujita *et al.* first noted that MTA3 is highly expressed in germinal-centre B cells. They went on to show that it physically associates with BCL-6. It thereby tethers the other Mi2–NuRD subunits to BCL-6 — an association that is essential for BCL-6 to repress its target genes, including *Blimp-1*. From a regulatory standpoint, these results highlight the fact that complexes that modify the packing of DNA can have cell-type-specific subunits that target the complexes to different genes in different cells.

Most unexpectedly, Fujita *et al.* also

found that introducing BCL-6 and MTA3 into plasma cells caused them to 'dedifferentiate' towards a B-cell state. The cells began to express numerous protein 'markers' of B cells. Conversely, they no longer expressed Blimp-1 or XBP1 (a key regulator of the ability of plasma cells to secrete antibodies⁹). Although this dedifferentiation is surprising, the potential for such plasticity is in fact built into double-negative regulatory networks¹.

An important caveat is that the dedifferentiation was accomplished with malignant plasma cells derived from a patient with multiple myeloma. The ability of BCL-6 and MTA3 to push these cells towards the B-cell state may have been helped by the fact that the cells are, abnormally, still dividing¹⁰. This is because they express *c-myc*, owing to a chromosomal translocation — an abnormal event in which segments of different chromosomes are swapped about, potentially affecting the expression of any genes involved, and often contributing to cancer. It will be interesting to conduct studies in which BCL-6 and MTA3 are overexpressed at various stages of normal plasma-cell differentiation, to see whether the same dedifferentiation results.

Turning a vice into a virtue, however, these experiments may help to explain some puzzling aspects of human lymphoid cancers. A cancer generally derives from one abnormal cell, which multiplies to produce a clone of malignant cells. In multiple myeloma, however, not all the cells in the so-called clonal population are alike: although most resemble plasma cells, a few lack plasma-cell proteins and instead express markers of B cells¹¹. This minor population of cells may be more proliferative than the rest, and there is evidence that they form tumours more readily in immunodeficient mice¹². This leads to the somewhat controversial notion that such cells are myeloma 'stem cells'^{11,12} — a pool of cells that can both multiply themselves and produce different cell types (in this case, the cells that resemble plasma cells, perhaps). Meanwhile, the cells that resemble plasma cells might be less proliferative, but they might also be long-lived and accumulate because of survival signals that they receive from the bone-marrow microenvironment in which they grow.

In light of Fujita and colleagues' findings, it is conceivable that the myeloma 'stem cells' are produced through the dedifferentiation of the cells that resemble plasma cells back into the germinal-centre B-cell state. It will be fascinating to see whether the proliferation and 'tumorigenicity' of myeloma cells is affected when the balance between BCL-6 and Blimp-1 is experimentally perturbed.

The new results might also see use in the clinic. The importance of BCL-6 in human lymphomas is already known: it is the most frequently translocated gene in non-

Hodgkin's lymphomas¹³, and is characteristically expressed by certain lymphoma subgroups¹⁴. Manipulating the BCL-6–Blimp-1 circuit might be useful in treating lymphomas, given that inhibiting BCL-6 activity in cultured lymphoma cells causes them to differentiate into plasma cells and stop dividing⁴. And now that MTA3 is known to be important to BCL-6 function, another window of opportunity has opened: next, we need to identify small-molecule inhibitors of this interaction. Perhaps the plasticity inherent in a double-negative regulatory circuit can be harnessed to therapeutic effect. ■

Louis M. Staudt is in the Lymphoid Malignancies Section, Metabolism Branch, Center for Cancer Research, National Cancer Institute, National

Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: lstaudt@mail.nih.gov

1. Monod, J. & Jacob, F. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961).
2. Fujita, N. *et al. Cell* **119**, 75–86 (2004).
3. Shaffer, A. L. *et al. Immunity* **17**, 51–62 (2002).
4. Shaffer, A. L. *et al. Immunity* **13**, 199–212 (2000).
5. Tunyaplin, C. *et al. J. Immunol.* **173**, 1158–1165 (2004).
6. Lin, Y., Wong, K. & Calame, K. *Science* **276**, 596–599 (1997).
7. Niu, H., Ye, B. H. & Dalla-Favera, R. *Genes Dev.* **12**, 1953–1961 (1998).
8. Allman, D. *et al. Blood* **87**, 5257–5268 (1996).
9. Shaffer, A. L. *et al. Immunity* **21**, 81–93 (2004).
10. Hollis, G. F., Gazdar, A. F., Bertness, V. & Kirsch, I. R. *Mol. Cell. Biol.* **8**, 124–129 (1988).
11. Szczypek, A. J. *et al. Blood* **92**, 2844–2855 (1998).
12. Matsui, W. *et al. Blood* **103**, 2332–2336 (2004).
13. Pasqualucci, L. *et al. Leuk. Lymphoma* **44** (Suppl. 3), S5–S12 (2003).
14. Shaffer, A. L., Rosenwald, A. & Staudt, L. M. *Nature Rev. Immunol.* **2**, 920–932 (2002).

Climate

Wider connections for El Niño

William J. Randel

Data from Europe in 1940–42, and simulations of severe El Niño events, suggest that the effects of such events can be unexpectedly far-reaching. The stratosphere could be a key player in this behaviour.

Climate variations associated with the El Niño/Southern Oscillation (ENSO) cause droughts, floods and extreme temperatures over much of the tropics, and the effects are often felt in North America. Brönnimann *et al.*¹ (page 971 of this issue) now show that climate anomalies associated with the large El Niño of 1940–42 extended to much more of the Northern Hemisphere, including Europe. The atmospheric circulation anomalies also reached into the stratosphere, influencing the ozone layer. Similar results from climate-model simulations likewise suggest that the effects of extreme ENSO events may be more far-reaching than previously thought.

El Niño is characterized by a warming of the surface water in the eastern Pacific Ocean that occurs irregularly every 3–7 years. The warming is caused by anomalous ocean circulation in the tropical Pacific, which is in turn linked to systematic changes in winds and storms throughout the Pacific region. The associated variations in atmospheric circulation were originally termed the Southern Oscillation, hence the designation ENSO for the coupled ocean–atmosphere cycles².

The tropical Pacific storms associated with ENSO have their far-reaching effects through the generation of planetary-scale atmospheric waves. These result in coherent climate anomalies — 'teleconnections' — over distant regions of the globe, which are experienced locally as persistent wet or dry periods, or systematic changes in storms or hurricanes. Teleconnections for ENSO include droughts or floods over parts of Africa, South Asia, Australia and South



Figure 1 Europe, winter 1940 — a Pacific link?

America. They also extend to high latitudes, in particular influencing temperatures and precipitation over North America by the so-called Pacific–North American teleconnection pattern. But most previous studies have suggested that ENSO has a relatively

small (or highly variable) impact on European climate³.

Enhanced planetary waves during ENSO events also propagate vertically and influence winter circulation in the stratosphere, although the effects are typically small compared with other sources of stratospheric variability. Systematic searches for ENSO effects in the stratosphere have shown only weak or variable signals⁴.

Brönnimann *et al.*¹ began by analysing some of the earliest measurements of atmospheric ozone, extending back to the 1930s at several European stations⁵. The data show unusually high and persistent values between 1940 and 1942, although the observations were never explained at the time. This wartime period also saw the first systematic (but sparse) balloon-borne measurements of the upper atmosphere. These measurements have been used in statistical 'reconstructions' of meteorological conditions in the upper atmosphere, and detailed comparisons with the ozone measurements show that the period of anomalously high ozone was associated with persistent meteorological anomalies in the lower stratosphere during winter. In particular, the polar stratosphere was relatively warm and stratospheric winds were relatively weak, characteristics that are typically associated with high ozone values.

During 1940–42, exceptional climate conditions occurred around the globe, including a series of intensely cold European winters (Fig. 1) that had a significant influence on events during the Second World War. A prolonged ENSO event occurred during 1939–42, and the coincidence of these anomalies suggests a connection between ENSO, European climate and the stratosphere. Brönnimann *et al.* have isolated the hemispheric-scale climate signals for this period, and find features expected for ENSO, including the Pacific–North American teleconnection pattern. But the results also highlight a large response over the Atlantic–European region, and an extension into the stratosphere (hence influencing ozone, as observed).

The anomalies in European climate and stratospheric circulation during the 1940–42 El Niño could be a chance occurrence, but Brönnimann *et al.* show that similar patterns occur in model simulations of extreme ENSO events. They analysed the largest ENSO events that occur in a 650-year model simulation, which includes coupling between atmosphere and ocean, and produces ENSO events as a part of its natural variability. An ensemble of the stronger ENSO events in the model shows global teleconnections that strongly resemble the 1940–42 observations, with strong surface signals in the Atlantic–European region that extend into the lower stratosphere.

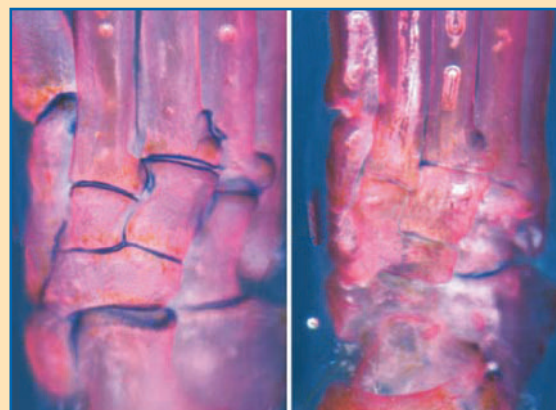
Both the spatial and temporal characteristics of the model events are a reasonable

Physiology

Joint approach

These photographs show what happens when a certain protein is inactivated in the ankle joints of developing mice. Particular joints completely fail to form. Elsewhere in the body, joints do form, but they are not maintained. Both conditions are evident from the lack of blue staining in the right-hand image. The photographs appear in a paper by David M. Kingsley and colleagues (*PLoS Biol.* **2**, e355; 2004), and the findings bear on our understanding of human osteoarthritis.

In healthy joints, where two bones meet, the ends of the bones are covered in cartilage, helping to reduce friction. In people with osteoarthritis, however, the cartilage wears away, so that the underlying bones are exposed and rub together painfully. To



investigate the molecular mechanisms involved, Kingsley and colleagues built on the knowledge that certain proteins of the bone morphogenetic protein (BMP) family are involved in joint formation. Using a complex genetic system, they generated mice in which a BMP receptor, the BMPRIa protein, is selectively inactivated in the joints.

In these mice, some joints did not form at all; others did form, but the cartilage

gradually wore away, producing the physical and behavioural characteristics of osteoarthritis. The authors suggest that signalling pathways activated by BMPRIa are needed to maintain the production of components of the cartilage extracellular matrix. And they propose that mutations in BMPRIa might account for some of the genetic variation that is known to contribute to human osteoarthritis. **Amanda Tromans**

match for the observed anomalies. Although ozone does not vary in this model, the stratospheric meteorology is consistent with large ozone anomalies. The overall agreement between the model and observations suggests that the 1940–42 climate anomalies corresponded to a recurring, extreme state of the climate system that was associated with El Niño and encompassed the troposphere and stratosphere.

So, might the stratosphere have an active role in extreme ENSO dynamics? Recent studies have highlighted a dynamical coupling between the stratosphere and troposphere during winter, with stratospheric pressure and wind anomalies at middle and high latitudes often being linked with surface anomalies of the same sign⁶. This coupling sometimes occurs in the form of downward-propagating anomalies, suggesting that the stratosphere can influence weather patterns at the surface^{7,8}. The amplitudes of the surface-pressure anomalies are often largest in the Atlantic region, where the north–south see-saw patterns contribute to a teleconnection structure known as the North Atlantic Oscillation, which in turn is strongly coupled to European climate⁹. Thus, a plausible connection is that large ENSO events affect winter stratospheric circulation at high latitudes, these anomalies in turn being communicated to the surface and having an especially strong influence in the Atlantic–European region.

This apparently cogent picture is marred by the absence of similar European and stratospheric anomalies for the large ENSO events of 1982–83 and 1997–98. This is not unexpected, however, given that the sample is so small. Each ENSO event is different in detail, and other factors are known to influence stratospheric circulation (such as the eruption of the El Chichon volcano in Mexico, in 1982). So although Brönnimann and colleagues' study is suggestive of a consistent chain of connections, the large variability between the handful of observed large ENSO events highlights the complexity of the troposphere–stratosphere climate system. ■

William J. Randel is in the Atmospheric Chemistry Division, National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307, USA.
e-mail: randel@ucar.edu

1. Brönnimann, S. *et al.* *Nature* **431**, 971–974 (2004).
2. Philander, S. G. H. *El Niño, La Niña and the Southern Oscillation* (Academic, San Diego, 1990).
3. Greatbatch, R. J., Lu, J. & Peterson, K. A. *Geophys. Res. Lett.* **31**, L02208 doi:10.1029/2003GL018542 (2004).
4. Hamilton, K. J. *Atmos. Sci.* **50**, 3468–3473 (1993).
5. Brönnimann, S., Luterbacher, J., Staehelin, J. & Svendsby, T. M. *Geophys. Res. Lett.* **31**, L08101 doi:10.1029/2004GL019611 (2004).
6. Thompson, D. W. J. & Wallace, J. M. *Geophys. Res. Lett.* **25**, 1297–1300 (1998).
7. Baldwin, M. P. & Dunkerton, T. J. *J. Geophys. Res.* **104**, 30937–30946 (1999).
8. Baldwin, M. P. & Dunkerton, T. J. *Science* **244**, 581–584 (2001).
9. Hurrell, J. W., Kushnir, Y., Ottensen, G. & Visbeck, M. *The North Atlantic Oscillation: Climate Significance and Environmental Impact* (Am. Geophys. Un., Washington DC, 2003).

Obituary

Maurice Wilkins (1916–2004)

The death on 5 October of Maurice Wilkins, following so soon on that of Francis Crick, inevitably invites reflections on the differences in character between these two close friends — who, with James Watson, shared the Nobel prize for the discovery that changed the face of biology, and indeed of medicine. The contrast in personality could scarcely have been greater. It was encapsulated by the television interviews that followed the announcement of the award in 1962. Crick, ebulliently expansive, threw off aphorisms. Wilkins, his coat collar turned up, eyed his interlocutor morosely. Invited to share the euphoria of the quest after scientific truth, he gave a characteristic response: “Oh, it’s all just fiddle, fiddle, fiddle.” Wilkins was never given to hyperbole.

In his research, Wilkins was neither a visionary nor a deep thinker like Crick. Rather, he was a meticulous, inventive and infinitely patient experimentalist. He loved instruments, and had an especial rapport with the workshop machinists. He adhered to the old string-and-sealing-wax tradition of experimental physics. He would prod about in dustbins for useful bits of metal, and visitors to the laboratory were sometimes startled to see a condom put to use as a gas box surrounding an X-ray camera. Wilkins had an almost tactile appreciation of interference and diffraction phenomena, and a pictorial perception of molecular structure. He liked models and mistrusted mathematical abstraction. His approach to structure, in the words of one collaborator, was “incredibly shrewd”.

Maurice Wilkins was born in New Zealand into an Anglo-Irish family of progressive Unitarian views. The family returned to England when he was six. He showed an early interest in microscopes and telescopes, and a precocious aptitude for building instruments. He graduated in physics from the University of Cambridge and then from Birmingham, where his PhD supervisor was the young John Randall. Thus began a close but never entirely congenial relationship.

Wilkins’s research was interrupted by the Second World War, when the department switched to radar research. Randall and his colleague Harry Boot developed the cavity magnetron, the all-important source of centimetric radar waves, while Wilkins was dispatched to E. O. Lawrence’s cyclotron laboratory in California to help fractionate uranium. When the war ended, Randall was



From DNA to the social responsibility of science

appointed professor of physics at the University of St Andrews in Scotland, where Wilkins joined him. Thence they moved to King’s College in London, where Randall first showed his remarkable — and, to the staid mandarins of the college, alarming — capacity for gathering funds and enlarging his department. He persuaded the Medical Research Council to establish a biophysics unit with Wilkins as deputy director.

Some three years earlier, Oswald Avery and his colleagues had demonstrated that DNA was the genetic material, but their work was largely ignored and, by many, disbelieved. But Wilkins thought otherwise and resolved to study its structure. After a period of skirmishing, using ultraviolet microscopy, with the disposition of DNA in cells, he began his studies on X-ray diffraction. He designed cameras and devices for pulling DNA fibres, and was fortunate in procuring some high-molecular-weight material from Rudolf Signer in Berne. By 1950 he had obtained diffraction pictures of unprecedented quality of what we now know as the A-form of DNA. These he showed at a meeting in Naples. In the audience was the young Jim Watson, who later wrote that Wilkins’s presentation “stood out like a beacon”, for it implied that the greatest prize — the structure of the gene — could be explicitly determined. When Watson asked to join Wilkins’s research group he received no encouragement. But from then on Watson’s course was clear, and quickly took him to Cambridge and his assignation with Francis Crick, so vividly described in *The Double Helix*.

Wilkins’s laborious progress towards

the structure of DNA was rudely disturbed by the arrival at King’s of Rosalind Franklin, to whom (behind Wilkins’s back) Randall had promised sole charge of the DNA problem. The disastrous consequences have been often related; the ensuing controversy caused Wilkins lasting pain, and most of all the jibe that he was against women.

The immediate upshot, at all events, was that he was cruelly elbowed out of his cherished project. He had surrendered Signer’s DNA to Franklin and she made good use of it. It was her superior B-form diffraction patterns that proved especially useful to Watson and Crick. Wilkins was too intimidated by Franklin’s formidable persona to begin model-building as he had planned, and by the time she left King’s it was too late: Crick and Watson had shot his fox.

We know now that Crick set out the reasons why Wilkins merited a share in the Nobel prize: Wilkins had initiated and almost single-handedly carried the DNA problem. He had “done numerous extensive, accurate and painstaking studies”. True, he had “worked rather slowly, but then hardly anybody else ha[d] done anything”. It was he who had recognized that the structure must be helical. Moreover, compelling as the DNA model appeared, it was still only a model; Watson in particular feared that it might after all be no more than a mirage. Wilkins and his colleagues at King’s spent the next years gathering experimental evidence that the structure was indeed essentially correct and refining its details.

After DNA, Wilkins began to think about neurobiology. There were notable publications on the structure of cell membranes, but gradually extramural interests supervened. Troubled by the misuse of science, Wilkins became the first president of the British Society for Social Responsibility in Science, and an active member of the Pugwash disarmament group. He organized an undergraduate course on the social aspects of science, which was hugely popular with the students and still thrives. He sat in on the course, a benevolent presence in the background, until a few months ago.

Amidst the furore about DNA, the merits of the participants and the apotheosis of Rosalind Franklin, Wilkins kept his peace until, only last year, he published his memoir, *The Third Man of the Double Helix*. It bears the stamp of the modesty and restraint that pervaded his whole career.

Walter Gratzer

Walter Gratzer is at The Randall Division, King’s College, New Hunt’s House, Guy’s Campus, St Thomas’ Street, London SE1 1UL, UK.
e-mail: wbg@helios.rai.ums.ac.uk

Biotechnology

Spot the single bacterium

Proc. Natl Acad. Sci. USA
doi:10.1073/pnas.0404806101 (2004)

Xiaojun Zhao and colleagues have devised a rapid detection system that can spot a single bacterium lurking in a sample. The system, which uses 'nanoparticles' loaded with fluorescent dye, produces a result within 20 minutes and could be used to check food and drinking water, make medical diagnoses, and combat bioterrorism.

The system is faster than other bacterial sensors because the particles contain thousands of dye molecules, for identifying bacteria through signal amplification, as opposed to the handful used in previous systems, say the authors. The particles are attached to specific antibodies that latch on to a target bacterium, producing a glow that can be seen under the microscope without the need to further amplify the signal.

Zhao and co-workers tested their design by spiking hamburger meat with the *Escherichia coli* strain O157:H7. As few as ten bacterial cells can cause illness, and this strain has caused several fatal outbreaks in Britain and elsewhere in recent years. The researchers could identify just one or two bacteria in a beef sample. What's more, by including a range of antibodies and dyes in the mix, they believe that the particles could be used to detect several different bacterial species in a single test.

Michael Hopkin

Cell biology

A question of size

Genes Dev. doi:10.1101/gad.1228804 (2004)

Biologists have long puzzled over how growing cells decide when they are big enough to divide. Paul Jorgensen *et al.* have found that, for budding yeast, this decision is heavily influenced by the rate of production of ribosomes — the cellular protein-making machines.

The authors had previously identified mutations in two genes, *SFP1* and *SCH9*, that cause cells to become particularly small. Now, Jorgensen *et al.* show that the Sfp1 and Sch9 proteins dictate the production of ribosomes. They find that concentrations of Sch9 in the cell and the movement of Sfp1 into the nucleus are tightly controlled by the quality of nutrients available to the cell. When concentrations of nutrients are high, Sfp1 and Sch9 are active and cells delay division until they have grown large and assembled a copious number of ribosomes. When nutrients are scarce, the activity of Sfp1 and Sch9 is low, causing the cell to compensate for the nutrient shortage by reducing the size and ribosome requirement for cell division.

Helen Pearson

Musical acoustics

Bach in tune with the times

J. Acoust. Soc. Am. **116**, 2416–2426 (2004)

You don't need to retune your piano to play Bach properly. Even though the title of his *Well-Tempered Clavier* suggests that the pieces were written for the 'well-tempered' method of tuning commonly used during the Baroque period, Michael F. Page concludes that the 'equal temperament' tuning method used for keyboards today is quite suitable for performing these works.

There is no perfect way of tuning a piano so that all the consonant musical intervals in an octave correspond to simple frequency ratios, and various compromises have been adopted at different stages in the history of music. Equal temperament imposes a constant spacing between each half-tone of the musical scale. The Baroque era was a time of change and experimentation in tuning systems, and the system most appropriate to Bach's 'well-tempered' pieces — perhaps the Werckmeister III tuning popular at the time — has been hotly contested.

Page now shows that, by taking into account not only the mathematical 'purity'



of intervals but also human perceptual responses, equal temperament generally works better than three other historical systems, including Werckmeister III. When the analysis is applied specifically to three preludes from the *Well-Tempered Clavier*, weighting each interval by the number of times it appears, equal- and well-tempered methods both perform strongly.

Philip Ball

Developmental biology

Junk no more?

Dev. Cell **7**, 597–606 (2004)

More than one-third of the human genome, previously thought to be non-functional, may in fact help to regulate gene expression, according to a study of mouse eggs and early embryos.

Anne E. Peaston *et al.* analysed sections of DNA that, when expressed, can make copies of themselves and jump into different positions within the genome. These chunks of DNA, known as retrotransposons, have generally been considered to be 'junk' because they seem to have no specific function.

The authors found, however, that certain retrotransposon sequences are surprisingly highly expressed in mouse eggs or early embryos — and that retrotransposon expression drove the expression of many genes. This seems to occur because the retrotransposons contain promoter sequences, DNA sections that trigger gene expression. Thus, they may provide a source of promoters that enable particular genes to be expressed specifically in eggs and at early stages of development. By inserting copies of themselves back into the genome, the retrotransposons might also, the authors suggest, generate gene mutations that promote genetic diversity.

Peaston *et al.* think that retrotransposon activity during early mammalian development might help to 'remodel' the

genome as necessary — affecting, for instance, the way in which different stretches of DNA are packed.

Mark Peplow

Biomedical materials

Eyeball glue

J. Am. Chem. Soc. **126**, 12744–12745 (2004)

Cataract removal by surgery could be made much less traumatic if a glue is used to seal the wound, Michel Wathier *et al.* report. Currently, the incision made in the eye's cornea to remove the clouded lens and replace it with a synthetic one is either sealed with nylon sutures or left to self-seal. Suturing can cause inflammation and other problems, and the stitches must subsequently be removed; self-sealing, meanwhile, runs the risk of fluid leakage from the eye and the development of infections.

The researchers have devised a biocompatible glue consisting of highly branched (dendritic) peptide molecules that can be chemically crosslinked at room temperature to form a viscoelastic, transparent hydrogel. Gelation occurs within minutes of mixing the peptide dendrimers and the crosslinking agents.

In tests on excised eyes, this type of repair prevented leakage of fluids pumped into the eye's anterior chamber at pressures more than three times greater than those that caused leakage from sutured incisions, and more than ten times greater than the normal pressures experienced in the eye.

Philip Ball

Leg feathers in an Early Cretaceous bird

This feature supports the idea that the evolution of flight involved a four-winged stage.

Here we describe a fossil of an enantiornithine bird from the Early Cretaceous period in China that has substantial plumage feathers attached to its upper leg (tibiotarsus). The discovery could be important in view of the relative length and aerodynamic features of these leg feathers compared with those of the small 'four-winged' gliding dinosaur *Microraptor*^{1,2} and of the earliest known bird, *Archaeopteryx*^{3,4}. They may be remnants of earlier long, aerodynamic leg feathers, in keeping with the hypothesis that birds went through a four-winged stage during the evolution of flight¹.

Some modern birds rely on structural features of their legs for assistance during flight — for example, the kittiwake (*Rissa tridactyla*) will spread out its webbed feet to provide an air brake, and the lappet-faced vulture (*Torgos tracheliotus*) stretches its legs and feet to their full extent in order to achieve a steep descent onto a kill^{5,6}.

Unlike modern birds, but like the basal dromaeosaur *Microraptor*, the legs of the new enantiornithine bird possess relatively long pennaceous feathers that are about half the length of the tibiotarsus; they are curved and situated mainly along the outer side of this bone (Fig. 1). Although the preservation of these feathers shows little evidence of asymmetry, their distinctive curvature may indicate a residual aerodynamic function^{3,5,7}.

In the four-winged dinosaur *Microraptor gui*, the extremely long leg feathers show unambiguous asymmetric vanes and curvature¹ that clearly had some aerodynamic function. The presence of leg feathers in *Archaeopteryx* has, after some controversy, been reconfirmed, although it is still not known whether they retained some aerodynamic function⁴. The leg feathers in the new enantiornithine are longer than those in *Archaeopteryx* when compared with the length of the tibiotarsus, and they appear to be more aerodynamic than the leg feathers in *Archaeopteryx*.

The tail feathers (rectrices) of modern birds contribute to lift during flight^{3,5,6,8}. Compared with what we know about the wing feathers (remiges) and skeleton, however, little is known about the role of the tail feathers of basal birds in the evolution of early avian flight⁸, even though many Mesozoic bird fossils have been discovered^{9–11}. The tail feathers of the new bird are short compared with those of modern birds, as is the case for many other Mesozoic birds, such as *Confuciusornis*, *Changchengornis*, *Protopteryx* and *Yixianornis*, although many of these specimens have a pair of long central tail feathers^{11–13}. It is possible that the short tail feathers

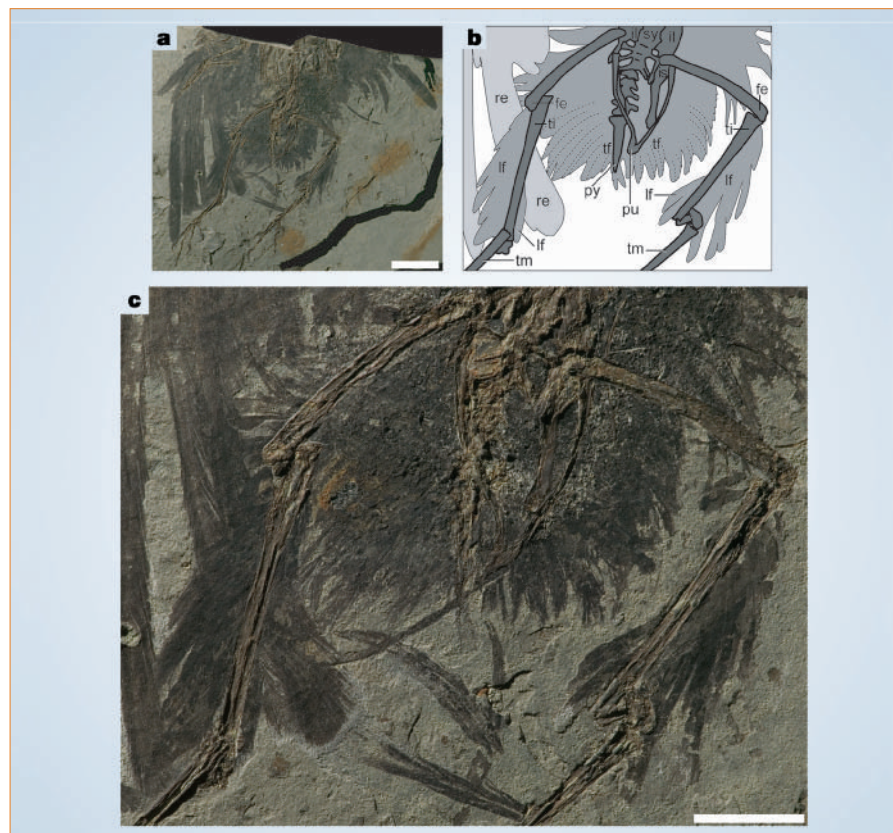


Figure 1 A fossil bird from the Early Cretaceous Yixian Formation at Jingangshan locality in Yixian, Liaoning Province, northeast China (IVPP Collection V 13939). Its caudally notched sternum, increased length of metacarpal III compared with II, and its reduced metatarsal IV indicate that it was an enantiornithine bird. **a**, Photograph of the specimen in a single slab, showing the association of the preserved skeletal elements, almost all of which are articulated, with feathers. **b**, **c**, Drawing (**b**) and close-up (**c**) of the specimen, which has long, aerodynamic feathers attached to its legs (lf) and weak, narrow and relatively short tail feathers (tf). Abbreviations in **b**: fe, femur; il, ilium; is, ischium; pu, pubis; py, pygostyle; re, remiges; sy, synsacrum; ti, tibiotarsus; tm, tarsometatarsus. Scale bars: **a**, 2 cm; **b**, 1 cm.

could be the result of moulting, although more evidence would be needed to corroborate this. If, however, the short tail feathers of the new bird and of other basal taxa are correctly recognized, then they might represent a primitive feature.

The flight ability of these early birds has been generally acknowledged^{3,9–11}, although it was probably not as strong as it is in modern birds. The weak aerodynamic features of the leg feathers of the new bird could have helped the poorly developed tail in manoeuvring during flight, rather as the modern razorbill (*Alca torda*) uses its large webbed feet to supplement its small tail during slow flight⁶.

Phylogenetically, enantiornithines are a more advanced group of birds than *Archaeopteryx*. The hindlimb wing of *Microraptor* was probably a primitive form that was later reduced in most birds. The leg feathers of *Archaeopteryx* and the new enantiornithine fossil may represent vestigial examples of a feature that made an important contribution to flight in early birds.

Fucheng Zhang, Zhonghe Zhou

Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China
e-mail: fuchengzhang@yeah.net

1. Xu, X. *et al.* *Nature* **421**, 335–340 (2003).
2. Prum, R. O. *Nature* **421**, 323–324 (2003).
3. Feduccia, A. *The Origin and Evolution of Birds* (Yale Univ. Press, New Haven & London, 1999).
4. Christiansen, P. & Bonde, N. *Comptes Rendus Palevol* **3**, 99–118 (2004).
5. Norberg, U. M. *Vertebrate Flight: Mechanics, Physiology, Morphology, Ecology and Evolution* (Springer, Berlin, 1990).
6. Pennycuik, C. *Animal Flight* (Arnold, London, 1972).
7. Norberg, R. A. in *The Beginnings of Birds* (eds Hecht, M. K. *et al.*) 303–318 (Freunde Jura-Museum, Eichstätt, 1985).
8. Gates, S. M. in *New Perspectives on the Origin and Early Evolution of Birds* (eds Gauthier, J. & Gall, L. F.) 237–254 (Peabody Mus. Nat. Hist., Yale Univ., New Haven, 2001).
9. Padian, K. & Chiappe, L. M. *Biol. Rev.* **73**, 1–42 (1998).
10. Witmer, L. M. in *Mesozoic Birds: Above the Heads of Dinosaurs* (eds Chiappe, L. M. & Witmer, L. M.) 3–30 (California Univ. Press, Berkeley, 2002).
11. Zhou, Z., Barrett, P. M. & Hilton, J. *Nature* **421**, 807–814 (2003).
12. Chiappe, L. M., Ji, S., Ji, Q. & Norell, M. A. *Bull. Am. Mus. Nat. Hist.* **242**, 3–89 (1999).
13. Chang, M., Chen, P., Wang, Y.-Q. & Wang, Y. (eds) *The Jehol Biota* (Shanghai Sci. Technol. Publ., Shanghai, 2003).

Competing financial interests: declared none.

Shotgun sequence assembly and recent segmental duplications within the human genome

Xinwei She¹, Zhaoshi Jiang¹, Royden A. Clark², Ge Liu², Ze Cheng¹, Eray Tuzun¹, Deanna M. Church³, Granger Sutton⁴, Aaron L. Halpern⁵ & Evan E. Eichler¹

¹Department of Genome Sciences, University of Washington School of Medicine, 1705 NE Pacific Street, Seattle, Washington 98195, USA

²Department of Genetics, Case Western Reserve University, Cleveland, Ohio 44106, USA

³National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, Maryland 20894, USA

⁴Applied Biosystems, 45 West Gude Drive, and ⁵The Center for the Advancement of Genomics, 1901 Research Boulevard, Suite 600, Rockville, Maryland 20850, USA

Complex eukaryotic genomes are now being sequenced at an accelerated pace primarily using whole-genome shotgun (WGS) sequence assembly approaches. WGS assembly was initially criticized because of its perceived inability to resolve repeat structures within genomes. Here, we quantify the effect of WGS sequence assembly on large, highly similar repeats by comparison of the segmental duplication content of two different human genome assemblies. Our analysis shows that large (>15 kilobases) and highly identical (>97%) duplications are not adequately resolved by WGS assembly. This leads to significant reduction in genome length and the loss of genes embedded within duplications. Comparable analyses of mouse genome assemblies confirm that strict WGS sequence assembly will oversimplify our understanding of mammalian genome structure and evolution; a hybrid strategy using a targeted clone-by-clone approach to resolve duplications is proposed.

The optimal method to generate assembled genomic sequence data for the scientific community has been a matter of considerable debate^{1,2}. Public efforts originally advocated strict clone-order-based approaches, citing logistical limitations, computational issues and an unknown genome structure as arguments against a WGS approach. In the case of the human genome, the clone-ordered approach involved the sequencing of large insert genomic clones (>100 kilobases (kb)) derived from a physical map generated before sequencing. This effectively reduced the genome project into a collection of local projects ($n = 45,000$) that could be subsequently assembled into a final genome sequence. The alternative WGS sequencing approach involved random sequencing of a large collection ($n = \sim 27,000,000$) of clones of various insert size as a single project. It assembled the genome sequence 'on the fly' based on sequence overlap and paired end-sequence linker information. Private initiatives demonstrated the efficacy of WGS sequence assembly to generate rapidly draft versions of eukaryotic genomes^{3,4}, although the inclusion of public clone-ordered data complicated interpretation of its power as a stand-alone strategy⁵. Since that time WGS assembly (WGSa)-based approaches have become widely adopted within the sequencing community and are now the predominant component of most publicly funded genome projects^{6,7}. Despite their general acceptance, the impact of such strategies on our understanding of genome biology is not well understood.

Recently, two independent assemblies of the human genome were released—one based largely on clone-ordered sequence (build34) and the other based on exclusive use of WGSa data. This landmark event provides the first opportunity to compare two distinct genome assembly approaches^{8,9}. It should be pointed out that both assemblies have matured over multiple rounds of reiteration. The current finished build34 benefited from several years of hand curation and experimental validation from a large number of genome annotators. Similarly, the WGSa was generated by an assembler that was enhanced after algorithmic improvements introduced during the Celera mouse assembly⁸. We present a

detailed study of the organization and structure of segmental duplications within these two assemblies. These results have important implications not only in directing and improving future genome assemblies, but, more importantly, in providing insight into how whole-genome sequence can be meaningfully interpreted by the biological community.

Segmental duplications and human assembly comparison

Both working-draft and WGS sequence assemblies have had difficulties resolving the structure of large, highly identical duplications^{10–13}. We analysed recent segmental duplications (>90% identity, >1 kb in length) using methods that were developed during the analysis of the human genome^{10,14}. Both experimental and *in silico* analyses initially suggested that 5–6% of human euchromatin is composed of segmental duplications^{9,15}. Precise determination of the organization and structure required a high-quality genome assembly. Within the finished build34 genome we identified 150.8 megabases (Mb; 5.3%) of segmental duplications (Table 1) of which 140.2 Mb could be confirmed using an assembly-independent strategy¹⁴, indicating that these were not artefacts of the build34 assembly (see Fig. 4b of ref. 9). Although this assembly represents a marked improvement from previous genome assemblies, gaps still remain particularly within duplication regions. Incremental improvements and increases in duplication content are expected. A more recent assembly of the human genome (build35), for example, captured an additional 2.0 Mb of duplicated sequence (Supplementary Table 1). A total of 98.7% of the duplication structure was identical between build34 and build35. In contrast to these, an independent analysis of the WGSa revealed a significantly reduced content of segmental duplications (60.3 Mb or 2.2% of the WGSa genome). The results of these duplication analyses including an overlay of WGSa on build34 are available in UCSC-browser format at <http://humanparalogy.gs.washington.edu>.

It had been predicted that duplications with the greatest degree of sequence identity would be the most difficult to resolve but, until

now, the threshold for this effect has been impossible to determine. For duplications with less than 95% identity there is a good correspondence between the two methods, although the WGSa shows fewer alignments at all bin sizes (Fig. 1a). At 95.5% a marked decline in the fraction of duplicated bases becomes apparent within the WGSa. As the sequence identity of duplications increases, the largest and most highly identical duplications disappear. We calculate that only ~9% (8.0 Mb out of the expected 94 Mb) of duplications whose sequence identity exceeds 97% are represented within the WGSa as duplications. Duplications that are virtually identical (>98%) appear to be completely absent within the assembly or take the form of apparent unique sequence composed of extremely short sequence alignments. The former corresponds to ~26 Mb of sequence (1,748 regions greater than 10 kb in length).

As expected, genes embedded within these segments are also conspicuously absent. We identified 67 genes that are partially deleted and another 36 genes that are completely absent from the Celera WGSa (Supplementary Table 2 and Supplementary Fig. 1). This set included rapidly evolving gene families such as nuclear-pore interacting protein (*NP1P*), sperm protein associated with nucleus (*SPANX*) and variable charge basic protein Y (*VCY*) gene families. In addition, cancer-related antigen markers (*GAGE*, *NYREN*), both survival of motor neuron genes (*SMN1* and *SMN2*) and several important immune-related genes—interleukin 27 (*IL27*), neutrophil cytosolic factor 1 (*NCF1*) and epithelial beta defensins (*DEF*)—were lost from the WGSa owing to their association with segmental duplications.

Next, we analysed the length distribution of duplication alignments between the two genome assemblies. Build34 duplications were distributed within a total of 28,728 pairwise alignments whose length averaged 9.2 kb. WGSa duplications were less frequent (20,818 alignments) and shorter (4.04 kb). An analysis of the sum of aligned bases as a function of alignment length showed a marked depletion of longer alignments (>15 kb) within the WGSa when compared with build34 (Fig. 1b). The greatest discrepancy occurred

among the largest alignments and pinpoints a failure of whole-genome assembly methods to traverse through such large, complicated repeat structures.

Large blocks of highly identical duplications are enriched near human centromeres and telomeres as well as specific focal regions within euchromatin. Not surprisingly, these regions are very poorly represented within the WGSa (Fig. 2; see also Supplementary Fig. 2). Such duplication regions are also frequently associated with genomic disease owing to non-allelic homologous rearrangement between intrachromosomal duplications. We examined the WGSa for five disease breakpoint regions (spinal muscular atrophy type I, Charcot–Marie–Tooth disease, velocardiofacial/DiGeorge syndrome, Prader–Willi syndrome and William’s syndrome). Between 71–97% of the sequence corresponding to these large segmental duplications was absent (Supplementary Fig. 1). It follows that strict dependence on a WGSa approach would severely oversimplify the architecture of our genome and limit an understanding of the molecular aetiology of such diseases.

As a final assessment of segmental duplications between the assemblies, we analysed 19 duplicons whose copy number and distribution within the human genome had been experimentally validated by fluorescence *in situ* hybridization and/or hybridization data. We mapped 75 sequence tags corresponding to these 19 duplicons by BLAST sequence similarity searches against the two human genome assemblies (Supplementary Table 3). Within the finished build34 genome a total of 535 copies mapped to specific chromosome positions—in good agreement with experimental data ($n = \sim 580$ copies). Eleven mapped to positions within the unplaced or random sequence contigs. By comparison, only 240 discrete loci could be identified within the WGSa. Of these, 94 mapped to specific chromosomal positions, whereas the remainder localized to an unknown chromosome. We conclude that a minor fraction (~20%) of the duplicated sequence is correctly placed within the WGSa and that more than one-half of the duplications have been collapsed or lost.

Table 1 Comparison of segmental duplication within two human genome assemblies

Chromosome	Build34 assembly			WGSa		
	Length (bp)	Duplication (bp)	Fraction	Length (bp)	Duplication (bp)	Fraction
1	221,562,941	11,553,369	0.0521	209,662,503	2,629,537	0.0125
2	237,541,603	10,000,492	0.0421	223,960,456	2,034,342	0.0091
3	194,473,779	3,299,552	0.0170	189,481,828	2,626,848	0.0139
4	186,841,959	4,287,299	0.0229	180,981,699	2,899,296	0.0160
5	177,552,822	5,956,951	0.0336	170,281,266	1,351,471	0.0079
6	167,256,575	3,600,793	0.0215	161,428,330	1,660,626	0.0103
7	154,676,518	13,096,209	0.0847	144,247,908	5,496,523	0.0381
8	142,347,919	3,250,852	0.0228	136,878,554	1,013,574	0.0074
9	115,624,042	11,096,428	0.0960	104,630,165	1,826,794	0.0175
10	131,173,206	8,937,553	0.0681	122,948,635	3,379,280	0.0275
11	130,908,854	5,535,297	0.0423	126,253,176	4,007,704	0.0317
12	129,826,277	2,922,438	0.0225	125,900,476	2,050,906	0.0163
13	95,559,980	3,212,091	0.0336	92,484,206	1,953,930	0.0211
14	87,191,216	1,587,527	0.0182	84,198,821	951,062	0.0113
15	81,259,656	8,577,567	0.1056	74,059,970	2,599,187	0.0351
16	79,932,429	9,124,179	0.1141	66,369,068	994,790	0.0150
17	77,677,744	7,746,457	0.0997	73,627,628	3,657,946	0.0497
18	74,654,041	1,898,132	0.0254	71,253,215	370,517	0.0052
19	55,785,651	4,051,295	0.0726	51,679,110	2,835,683	0.0549
20	59,424,990	1,479,847	0.0249	57,238,069	963,199	0.0168
21	33,924,307	1,791,042	0.0528	31,584,736	410,918	0.0130
22	34,352,051	3,982,963	0.1159	31,357,605	1,590,197	0.0507
X	149,215,391	10,057,692	0.0674	121,809,144	2,105,297	0.0173
Y	24,649,555	12,745,541	0.5171	7,151,840	728,694	0.1019
Unplaced	2,592,022	980,700	0.3784	36,146,472	10,186,469	0.2818
Total	2,865,069,170	150,772,266	0.0530	2,695,614,880	60,324,790	0.0224

Segmental duplications (>90% sequence identity and >1 kb length) were calculated using the whole-genome assembly comparison method¹⁰ for the finished human genome assembly (July 2003) and the whole-genome shotgun sequence assembly (WGSa)⁸. Due to the fragmentation of duplications within the WGSa, duplicated bases were calculated without welding across gaps in the assembly. Totals do not include gaps or centromeric/acrocentric regions of chromosomes. Both assemblies were compared using exactly the same parameters. The unplaced chromosome contains the largest proportion of WGSa duplicated sequence—28.2% (10.2 Mb based on the analysis of WGSa). Of the 21.8 Mb that could be mapped back to build34, we found that 9.2 Mb (42.3%) corresponded to duplications within our segmental duplication database.

Segmental duplications and WGS chromosome length

The size of human euchromatin within build34 (2,865 Mb) is significantly larger than that predicted by the WGS (2,696 Mb). This size difference is not uniformly distributed among chromosomes (Table 1; see also Supplementary Fig. 3). Although some of this lost euchromatin, 170 Mb, has been attributed to reduced coverage of the sex chromosomes as a result of the male donor in the WGS⁸, differences in the length of the sex chromosomes would only account for 44.9 Mb of sequence. As human chromosomes vary considerably in duplication content, we sought to determine whether there was a correlation between chromosomes that carry large blocks of duplications with high sequence identity and reduced chromosomal length (Supplementary Fig. 3). There is a strong correlation ($r = 0.9$) between the reduction in chromosome length and the number of highly identical duplication bases (Table 1). Chromosome 16 is most notable in this regard. This chromosome is reduced by 17% in the WGS. It is also the autosome that has the largest fraction of highly identical segmental duplications (Table 1). We estimate that missing segmental

duplications contribute more than 50% to the reduced size of the WGS when compared with build34 (90 Mb out of the 170 Mb reduced size).

Implications

Our analysis clearly shows that strict WGS has limited capacity to resolve the structure of duplicated regions within genomes. Most of this effect, however, occurs among duplications that exceed >15 kb in length and show greater than 97% sequence identity. We predict that the largest, nearly identical duplications will be absent from WGS sequence assemblies. Clearly, different assembly algorithms^{16,17} may perform better or worse than the Celera assembler—with a trade-off between ability to separate repeats and robustness in the face of polymorphism or sequencing error—but these thresholds provide a useful benchmark for future genome assembly comparison. This study has several important ramifications.

First, estimates of genome-wide duplication content between species should be tempered by the underlying method of assembly. Apparent differences in content may be a consequence of differences in genome assembly rather than a true biological effect. This may explain why other complex genomes that have been sequenced primarily by WGS sequencing show reduced recent duplication content when compared with human^{13,18}. In such first-pass, whole-genome assemblies it is likely that duplications will be even more grossly underestimated. For example, we compared two mouse genome assemblies: one was assembled almost strictly by WGS sequencing (MGSCv.3.0) and the other was the most recent composite assembly (build33) where 57% of the mouse genome was assembled from large insert bacterial artificial chromosome (BAC) clones. The proportion of segmental duplication (>20 kb) increased by more than one order of magnitude between the two assemblies (Supplementary Table 4), where almost all of the increase (96%) was attributed to the incorporation of BAC-based sequence into the assembly.

Second, we can expect that euchromatin length will be underestimated on the basis of the misassembly of large, highly identical duplications. For the human genome sequence this effect accounts for more than 50% of the reduced genome length. It follows that organisms with greater duplication content will show greater reductions in size if WGS is the only method applied.

Third, genes embedded within segmental duplications will be concomitantly lost. A surprising finding was that 37 duplicated gene segments were not represented even once within the assembly (Supplementary Table 2). This may be because of fracturing of the assembly due to conflicting overlap patterns and mate-pair conflicts, leading to complete rejection of these regions. In the case of the sequenced human genome, this included the loss of rapidly evolving lineage-specific gene families and genes associated with immune response and germline development.

Most importantly, an oversimplified view (Fig. 2) of the human genome structure emerges with strict WGS. Regions enriched for duplications, such as pericentromeric and subtelomeric areas of chromosomes, are particularly under-represented (Supplementary Fig. 2). In addition, sites of recurrent chromosomal structural rearrangement associated with disease¹⁹ and breakpoints in conserved synteny essentially disappear as a result of WGS^{20,21}. In the absence of BAC-based sequence we will forfeit an understanding of heterochromatic–euchromatic transition regions, potential mechanisms of chromosomal evolution and the molecular aetiology/or origin of human genomic disease.

Hybrid strategy to sequence complex genomes

Although it is clear that the detailed clone-ordered approach is superior in the resolution of segmental duplications, it would be unrealistic to propose that the sequencing community should abandon WGS-based approaches. These are the most efficient

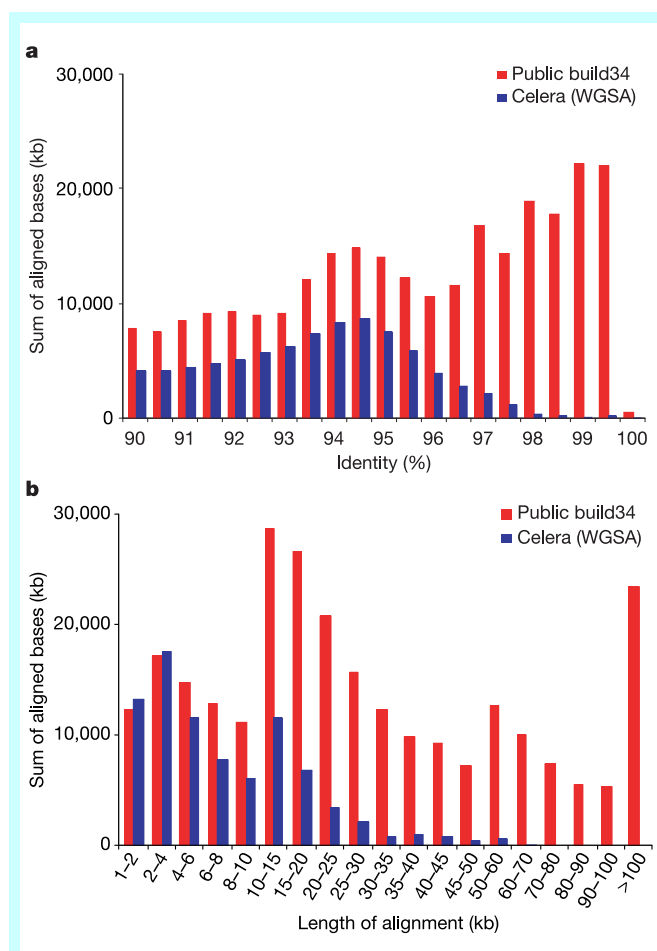


Figure 1 Sequence identity and alignment length of segmental duplications. **a**, **b**, All duplication alignments between 90–100% were categorized based on sequence identity (**a**) (0.5% bins) and the alignment length (**b**). The sum of aligned base pairs for each bin is compared between WGS and build34 human genome sequence assemblies. The proportion of WGS aligned base pairs begins to decline most rapidly as the sequence identity exceeds 96–97% and the length of the alignments exceeds 15 kb. Note that the reduction in WGS alignments below 96% is probably due to the fact that divergent duplications are frequently part of larger alignments where the degree of sequence identity is higher. As highly identical alignments are lost, the embedded, more divergent pairwise alignments are also eliminated from further consideration.

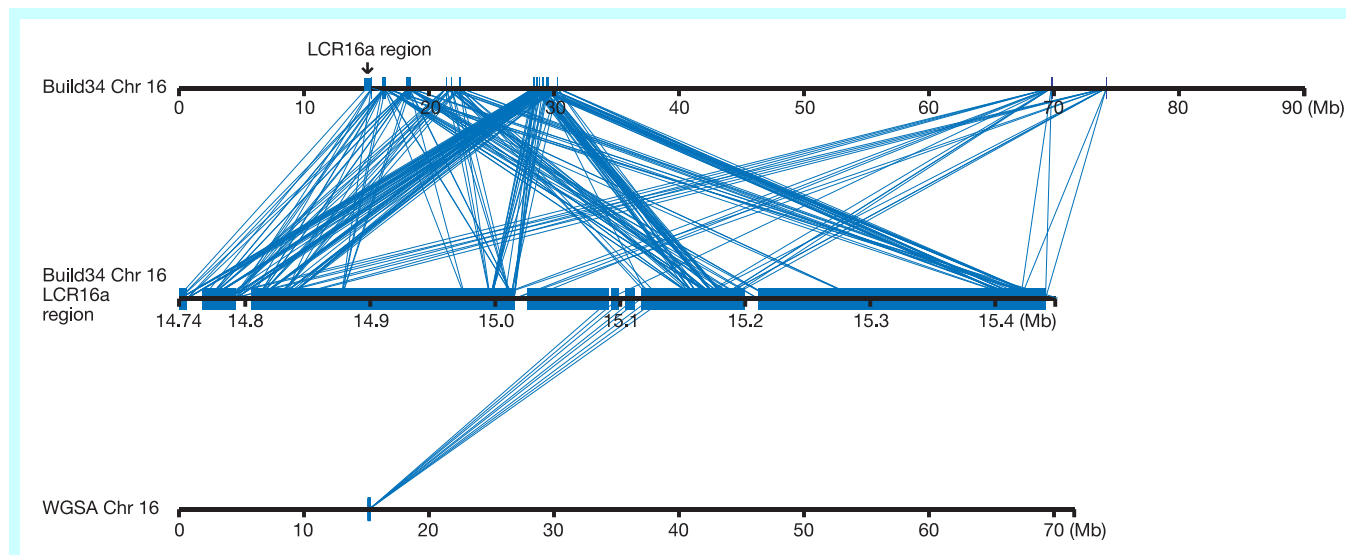


Figure 2 Distribution of LCR16a duplications in two assemblies. The pattern of duplication alignments for one 690-kb region of low-copy-repeat duplications on chromosome 16 is shown between the build34 and WGS human genome assemblies.

The entire region is duplicated to 28 distinct regions within build34 (locations have been experimentally verified) whereas only a small portion (46 kb) maps to a single location on WGS chromosome 16.

and cost-effective means of capturing the bulk of euchromatic sequence. Segmental duplications, however, should not be considered as an acceptable casualty of this process. In humans, duplicated regions show high transcriptional content²², are associated with disease¹⁹ and large-scale copy number polymorphisms²³, and have played an important role in the chromosomal evolution of mammalian genomes^{18,20}. Although the precise balance of clone-ordered sequencing and WGS sequencing during the assembly process has yet to be determined, the availability of two methods of genome assembly provides an important insight into this issue by refining the precise limitations of the WGS approach.

We propose a two-tier plan to ensure the resolution of such regions. During the first phase, WGS-based assemblies would be generated at sufficient depth (5–7-fold coverage) to provide an initial draft assembly of a genome. The same sequence reads could then be remapped to the assembly and analysed for regions of excessive divergence and excessive read depth as a means to detect sites of potential duplication¹⁴. Caution must be exercised to ensure that short sequence contigs are not completely excluded during this process, as those that do not originate from bacterial contamination often map to repetitive or duplicated portions of the genome. During the second phase, BACs corresponding to these regions of excess divergence and read depth would be selected based on BAC end sequence placement and submitted for further mapping/sequencing to establish long-range continuity across these regions. Sequence from these large-insert clones would be preferentially integrated into the WGS. Retrospectively, on the basis of the known human genome structure, we estimate that 94 Mb of genomic sequence within ~380 regions of the human genome would require BAC-based sequence. This would entail the pre-selection of approximately 3,000 BACs (2,300 BACs at fourfold coverage plus an additional 700 BACs that span transition regions). Low-level draft sequence and/or high-density fingerprinting would reduce this set to a minimal tiling path for higher quality sequence. In theory, WGS sequencing (~sixfold genome coverage) coupled with final clone-order-based sequencing of ~1,500 BAC clones would be sufficient to represent accurately the true architecture of the human genome. □

- Green, P. Against a whole-genome shotgun. *Genome Res.* **7**, 410–417 (1997).
- Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Waterston, R. H., Lander, E. S. & Sulston, J. E. More on the sequencing of the human genome. *Proc. Natl Acad. Sci. USA* **100**, 3022–3024 (2003) author reply 3025–3026.
- Rat Genome Sequencing Project Consortium, Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Mouse Genome Sequencing Consortium, Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Istrail, S. *et al.* Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl Acad. Sci. USA* **101**, 1916–1921 (2004).
- International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature* doi:10.1038/nature03001 (this issue).
- Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
- Cheung, J. *et al.* Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* **4**, R25 (2003).
- International Human Genome sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Tuzun, E., Bailey, J. A. & Eichler, E. E. Recent segmental duplications in the working draft assembly of the brown Norway rat. *Genome Res.* **14**, 493–506 (2004).
- Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
- Cheung, V. G. *et al.* Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**, 953–958 (2001).
- Huang, X., Wang, J., Aluru, S., Yang, S. P. & Hillier, L. PCAP: a whole-genome assembly program. *Genome Res.* **13**, 2164–2170 (2003).
- Batzoglou, S. *et al.* ARACHNE: a whole-genome shotgun assembler. *Genome Res.* **12**, 177–189 (2002).
- Bailey, J. A., Church, D. M., Ventura, M., Rocchi, M. & Eichler, E. E. Analysis of segmental duplications and genome assembly in the mouse. *Genome Res.* **14**, 789–801 (2004).
- Stankiewicz, P. & Lupski, J. R. Genomic architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
- Armengol, L., Pujana, M. A., Cheung, J., Scherer, S. W. & Estivill, X. Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. *Hum. Mol. Genet.* **12**, 2201–2208 (2003).
- Bailey, J. A., Baertsch, R., Kent, W. J., Haussler, D. & Eichler, E. E. Hotspots of mammalian chromosomal evolution. *Genome Biol.* **5**, R23 (2004).
- Hillier, L. W. *et al.* The DNA sequence of human chromosome 7. *Nature* **424**, 157–164 (2003).
- Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.E.E. (eee@gs.washington.edu).

Received 21 February; accepted 27 September 2004; doi:10.1038/nature03062.

1. Weber, J. L. & Myers, E. W. Human whole-genome shotgun sequencing. *Genome Res.* **7**, 401–409 (1997).

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

* A list of authors and their affiliations appears in the Supplementary Information

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

The Human Genome Project (HGP) was launched in 1990 with the goal of obtaining a highly accurate sequence of the vast majority of the euchromatic portion of the human genome. The initial work followed a two-pronged approach: (1) the mapping of the human and mouse genomes^{1–9} to allow the study of inherited disease and provide a crucial scaffold for genome assembly; and (2) the sequencing of organisms with smaller, simpler genomes^{10–14} to serve as a testbed for method development and assist in interpreting the human genome. With success along both paths, the sequencing of the human genome itself eventually became feasible. The International Human Genome Sequencing Consortium (IHGSC), an open collaboration involving twenty centres in six countries, was formed to carry out this component of the HGP.

In February 2001, the IHGSC¹⁵ and Celera Genomics¹⁶ each reported draft sequences providing a first overall view of the human genome. These sequences allowed systematic study of the human genome itself, including identification of genes, combinatorial architecture of proteins, regional differences in genome composition, distribution and history of transposable elements, distribution of polymorphism and relationship between genetic recombination and physical distance. Moreover, systematic knowledge of the human genome has enabled new tools and approaches that have markedly accelerated biomedical research.

Both draft sequences, however, had important shortcomings. The IHGSC sequence, for example, omitted ~10% of the euchromatic genome; it was interrupted by ~150,000 gaps; and the order and orientation of many segments within local regions had not been established. The IHGSC thus turned to the challenge of completing the sequence of the euchromatic genome. Operationally, a finished sequence was defined as having an error rate of, at most, one event per 10⁴ bases, and the goal for completion was coverage in finished sequence of at least 95% of the euchromatic genome, with the only gaps being those refractory to all available techniques¹⁷ (see <http://www.genome.gov/10000923>). The goal was challenging because the human genome is replete with such features as dispersed repeats and large segmental duplications, which greatly complicate the determination of genome structure and sequence. In fact, near-complete sequences have been obtained so far only for three multicellular organisms: the nematode¹³, mustard weed¹⁸ and the fruitfly¹⁹. These genomes are all roughly 30-fold smaller than the human genome and have much simpler structure.

We describe here the results of a multiyear effort by the IHGSC

towards the goal of a complete human sequence. The number of gaps has been reduced 400-fold to only 341, most of which are associated with segmental duplications and will require new methods for resolution. The assembled near-complete genome sequence has an error rate of only ~1 event per 100,000 bases; it contains 2.85 billion nucleotides and covers ~99% of the euchromatic genome. This paper describes the current genome sequence and the process used to produce it; examines the accuracy and completeness of the sequence; and illustrates biological analyses made possible by the sequence. We do not attempt here a comprehensive analysis of the contents of the human genome. An initial analysis was previously reported¹⁵ and a series of papers is being written describing the individual chromosomes^{17,20–30}, including annotation of genes and other features.

Current genome sequence

Finishing process

The process of converting the initial draft sequence into a near-complete sequence is referred to as ‘finishing’. It is a complex iterative process that proceeds simultaneously at multiple scales, ranging from single nucleotides to the integrity of whole chromosomes. The fundamental challenge is that genomic regions that are not well represented or readily resolved through random shotgun sequencing tend to be highly enriched in problematic sequences. Resolving such regions required the development of special approaches, which evolved substantially over time and varied among centres.

Broadly, the finishing process involved two distinct components: (1) producing finished maps, consisting of continuous and accurate paths of overlapping large-insert clones spanning the euchromatic region of each chromosome arm; and (2) producing finished clones, consisting of continuous and accurate nucleotide sequence across each large-insert clone. In practice, these two components were tightly intertwined in that progress in each often depended on results from the other. The components are described in Boxes 1 and 2. Further information about the finishing process and finishing standards can be found in the Supplementary Information (Note 1) and at <http://www.genome.gov/10000923>.

In total, we generated a shotgun sequence from 59,208 large-insert clones (total length ~5.84 gigabases (Gb)) and finished the sequence from 45,742 of these clones (total length ~3.67 Gb). The clones consisted primarily of bacterial artificial chromosomes

(BACs), but also included some P1-artificial chromosomes (PACs), yeast artificial chromosomes (YACs), fosmids and cosmids; they carried DNA from multiple anonymous sources¹⁵. We then chose a 'clone tiling path' of 26,720 overlapping clones across the genome, selected a 'sequence tiling path' of directly adjacent, non-overlapping segments from consecutive clones and concatenated these segments to create a near-complete genome sequence. Contributions of the IHGSC centres to this finishing phase are shown in Table 1.

Genome sequence

The human sequence reported here consists of 2,851,330,913 nucleotides, lying almost entirely within the euchromatic portion

of the genome (Table 2). It is interrupted by only 341 gaps, of which 33 gaps (totalling ~198 megabases (Mb)) reflect heterochromatin, which was not targeted by the HGP, and 308 gaps (totalling ~28 Mb) are euchromatic. The euchromatic genome is thus ~2.88 Gb and the overall human genome is ~3.08 Gb. The long-range continuity of the current genome sequence is high by various measures (Table 3). The N50 length is 38.5 Mb and the *N*-average length is 40.9 Mb; these values are ~1,000-fold larger than the size of a typical human gene. (The first statistic is the length *x* such that at least 50% of nucleotides lie in a continuous segment of length $\geq x$, whereas the second is the average length of the contiguous segment containing a randomly chosen nucleotide.) Focusing on individual

Box 1

Finishing the physical map

The hierarchical strategy used two kinds of genome maps as a foundation for producing finished sequence: sequence-tagged site (STS) maps⁶ and clone maps⁷. The first provided global landmarks by positioning tens of thousands of STSs through genetic mapping, radiation hybrid mapping and STS-content mapping. The second provided overlapping clones for sequencing and was obtained by comparing restriction-digest fingerprints of hundreds of thousands of BAC clones to create local contigs. The two maps were integrated by anchoring the contigs to the global landmarks.

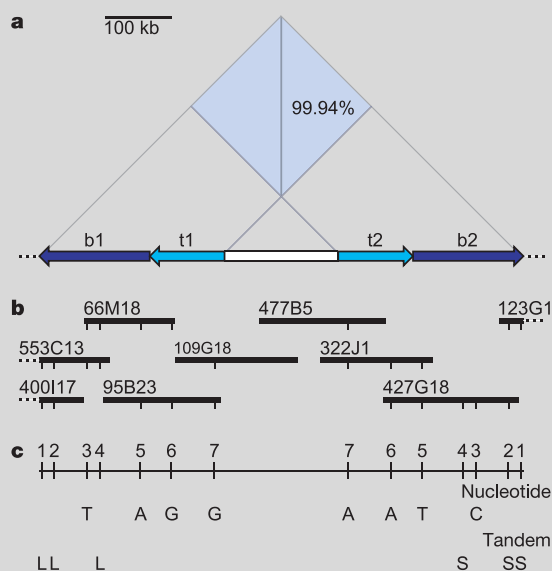
The initial random phase generated a physical map covering 96–98% of the euchromatic genome in ~1,000 anchored contigs separated by gaps with average size of ~100 kb⁷. Overlapping clones were chosen from this map to produce the draft sequence. The subsequent finishing phase involved verifying clone overlaps and closing gaps in the initial map. The finishing process for each chromosome was overseen by a designated centre and managed by a dedicated coordinator, who integrated and reviewed information from contributing centres.

Clone overlaps were verified by analysing finished sequence to confirm that the terminal sequence overlapped for ≥ 2 kb with $\geq 99.6\%$ identity. (Perfect identity was not expected in all cases, because the clones might derive from different haplotypes and thus differ at polymorphic sites.) A small number of overlaps ($n = 308$) falling below this threshold were accepted, because the sequencing centre presented additional evidence that the overlap was correct. For example, smaller overlaps might be confirmed by a partially sequenced clone spanning the overlap, or a region of high variation might be shown to be due to allelic variation. All these exceptions are recorded in electronic tags shown on the UCSC genome browser. Remaining overlaps were rejected and counted as new gaps in the map.

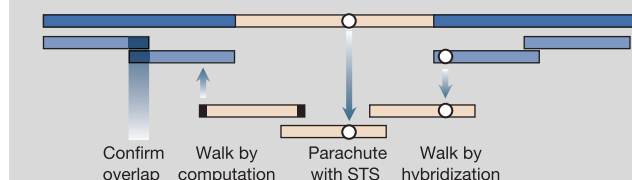
Closing gaps required identifying a tiling path of clones spanning the region. Methods used are summarized in Box 1 Fig. 1. The primary approach was iterative 'walking' from the ends of contigs: a stretch of terminal sequence from a terminal clone was used to identify a new clone extending the contig. The procedure was repeated until it either yielded a clone connecting to the neighbouring contig or reached a dead end. Walking used both experimental (hybridization to arrayed BAC libraries) and computational (comparison with a database of BAC end-sequences) methods to identify candidate clones. Also, it was

sometimes possible to 'parachute' into gaps by identifying sequences likely to lie within the gap (such as STSs, mRNAs and mouse sequences in regions of conserved synteny) and using them as hybridization probes for BACs.

Some gaps could not be closed, either because no new clone could be found despite screening of diverse and deep libraries (>30-fold physical coverage) or because a complex collection of clones was found whose relationship proved impossible to discern owing to the presence of extensive highly repetitive sequences or near-exact segmental duplication. Telomeric regions of chromosomes were obtained by using a specialized cloning system ('half-YACs vectors') in which successful propagation required that the human insert DNA provided telomeric function. By far, the most difficult regions of the genome were those containing near-exact segmental duplications. A particularly challenging example is shown in Box 1 Fig. 2.



Box 1 Figure 2 Illustration of a challenging region on chromosome Y. **a**, The sequence organization of the palindromic repeat P3 is represented by the horizontal bar at the base of the triangle. Arrows indicate orientation of sequences in the 283-kb arms of the palindrome (b1, t1 on left and t2, b2 on right). A non-repeated 170-kb 'spacer' (white) separates the arms. Above the horizontal bar, each dot represents a perfect match of 500 bp. The near identity between the arms (99.94%) appears as a vertical line of dots, highlighted by the blue diamond. **b**, Tiling path of BACs from RPCI-11 library. **c**, Sequence differences, numbered 1–7, between arms of P3. Differences 3, 5, 6 and 7 are single nucleotide differences, whereas 1, 2, and 4 are differences in the lengths of simple tandem repeats (microsatellites); L, long variant; S, short variant. These differences allowed assignment of BACs to the correct arm of P3.



Box 1 Figure 1 Simplified flowchart for finishing the physical map.

Table 1 **Bases sequenced in Build 35**

Centre	Finished sequence totals (kb)	Finished sequence in Build 35 (kb)
SC	919,388	849,650
WUGSC	645,062	583,032
WIBR	562,096	373,760
JGI/SHGC	485,085	313,988
BCM	320,735	280,963
RIKEN	155,769	112,047
UWGC	145,745	105,573
GS	99,970	78,467
GTC	45,710	32,972
UWMSC	39,227	28,367
Keio	44,905	20,780
IMB	73,677	20,053
Beijing	38,079	17,114
MPIMG	9,838	5,673
GBF	8,325	5,547
UOKNOR	18,657	5,311
TIGR	10,390	3,054
CGM	2,768	1,766
SDSTDC	7,792	1,403
UTSW	8,555	196
Other	30,745	11,621
Total	3,672,516	2,851,336

Columns indicate total finished sequence deposited in public databases (including overlaps and alternative alleles) and finished sequence incorporated into Build 35 (consisting of those clones chosen for inclusion in the tiling path by the individual chromosome coordinators). The total includes finished sequence completed at the time of the draft sequence^{15,17,20} and subsequently. Some clones were sequenced to draft coverage by one centre, then finished by another. Abbreviations: SI, Wellcome Trust Sanger Institute (Hinxton, UK); WUGSC, Washington University Genome Sequencing Center (St Louis, USA); JGI, US Department of Energy Joint Genome Institute (Walnut Creek, USA); WIBR, Whitehead Institute for Biomedical Research (Cambridge, USA); BCM, Baylor College of Medicine Human Genome Sequencing Center (Houston, USA); RIKEN, RIKEN Genomic Sciences Center (Yokohama, Japan); UWGC, University of Washington Genome Center (Seattle, USA); GS, Genoscope (Evry, France); Keio, Keio University School of Medicine (Tokyo, Japan); GTC, Genome Therapeutics Corporation (Waltham, USA); UWMSC, Institute for Systems Biology Multimegabase Sequencing Center (Seattle, USA); IMB, Institute of Molecular Biology (Jena, Germany); Beijing, Beijing Genomics Institute, Beijing, China; UOKNOR, University of Oklahoma's Advanced Center for Genome Technology (Norman, USA); SHGC, Stanford Human Genome Center (Stanford, USA); MPIMG, Max Planck Institute for Molecular Genetics (Berlin, Germany); GBF, German Research Center for Biotechnology (Braunschweig, Germany); TIGR, The Institute for Genome Research (Rockville, USA); CGM, Center for Genetics in Medicine (Perkin Elmer/Washington Univ.); SDSTDC, Stanford DNA Sequencing and Technology Development Center (Stanford, USA); UTSW, University of Texas, Southwestern Medical Center (Dallas, USA). 'Other' includes clones from groups that deposited <1 Mb of finished sequence that comprises Build 35. The total includes 5,387 bp, distributed across nine chromosomes, that were ambiguous (scored as 'N') and therefore not counted in the total figure used in the text.

chromosome arms, the N50 length exceeds half the length of the arm in three-quarters of cases (Table 3).

The sequence is denoted as NCBI Human Build 35 (May 2004), with the individual chromosomes having accession numbers NC000001 to NC000024 (see Supplementary Information Note 3 concerning additional sequence data). The analyses reported here were performed on Build 35 or, in a few cases, its immediate predecessor, Build 34 (which differed only slightly). The poster accompanying this paper displays the 24 human chromosomes, together with various biological annotations. These include GC content, repeat content, segmental duplications, protein-coding genes, sequence similarity and syntenic conservation with mouse, sequence similarity with the pufferfish, and density of single-nucleotide polymorphisms in the human genome. Many additional annotations can be found on public genome browsers (<http://genome.ucsc.edu/>; <http://www.ensembl.org/>; <http://www.ncbi.nlm.nih.gov/genome/guide/human/>), which are regularly updated.

Comparison with draft sequence

The near-complete sequence is a great improvement over the earlier draft sequence. It has substantially fewer gaps (341 versus 147,821) and greater continuity (38,500 kilobases (kb) versus 81 kb for N50 contig size), reflecting an overall improvement of ~475-fold. The draft sequence contained regions in which the local order and orientation were unknown; these have now been resolved. The case of chromosome 7 is illustrated in Fig. 1. Additionally, the draft sequence contained substantial artefactual duplication, including local events caused by errors in merging some adjacent BAC-based sequences, made by the first-generation global assembly program, and global events caused by contamination of shotgun assemblies of some BACs with data from other clones. These artefacts have now been eliminated.

Accuracy and completeness

Because the human genome sequence is intended to serve as a

Table 2 **Finished sequence and gaps, HGSC Build 35**

Chr	Total finished sequence* (kb)	Euchromatic gaps†		Heterochromatic gaps‡		Estimate of total gap size§ (kb)	Unfinished clones	
		Number	Est. size (kb)	Number	Est. size (kb)		Number	Est. size (kb)
1	222,828	32	1,605	2	19,510	21,115	17	850
2	237,503	20	2,512	1	2,900	5,412	0	0
3	194,636	5	1,935	1	1,500	3,435	0	0
4	187,161	14	1,250	1	3,000	4,250	0	0
5	177,703	5	92	1	340	432	0	0
6	167,318	10	658	1	2,300	2,958	0	0
7	154,759	11	869	1	4,630	5,499	0	0
8	142,613	9	662	1	2,190	2,852	0	0
9	117,781	40	1,955	2	18,000	19,955	12	600
10	131,614	12	1,020	1	2,515	3,535	8	400
11	131,131	7	322	1	4,760	5,082	0	0
12	130,259	8	795	1	4,300	5,095	0	0
13	95,560	6	715	2	17,200	17,915	0	0
14	88,291	1	8	2	17,220	17,228	0	0
15	81,342	10	737	2	18,260	18,997	0	0
16	78,885	4	143	2	10,000	10,143	0	0
17	77,800	9	875	1	7,500	8,375	0	0
18	74,656	3	97	1	1,368	1,465	0	0
19	55,786	5	5,015	1	340	5,355	0	0
20	59,505	4	1,157	1	1,766	2,923	0	0
21	34,170	3	53	2	11,620	11,673	0	0
22	34,765	11	460	2	14,330	14,790	0	0
X	150,394	12	750	1	3,000	3,750	14	700
Y	24,872	9	1,480	2	31,618	33,098	7	350
Total	2,851,331	250	25,165	33	200,167	225,332	58	2,900

* The total length of tiling paths including only finished bases of clones in Build 35. Roughly 2.19 Mb of sequence on chromosome Y was derived directly from the equivalent pseudoautosomal region on chromosome X.

† Defined as gaps in euchromatic regions, including junctions with heterochromatic/centromeric sequences, for which no clone was available (see text).

‡ Defined here as gaps in heterochromatic regions (see text and Supplementary Note 2 on heterochromatic sequence). Separate gaps were counted for centromeres and pericentromeric heterochromatin, even when the two were contiguous. Centromere sizes were taken from ref. 62 or in some cases provided directly by the sequencing centres (see Supplementary Note 2). Acrocentric sizes are based on centromere ratios from ref. 63. The sizes of large heterochromatic gaps are typically difficult to estimate accurately owing to their repeat structure and polymorphic nature^{62,64}. Other regions might arguably be called heterochromatin (for example, the pericentromeric regions of chromosomes 19 and 3 and a ~400-kb gap on the Y chromosome⁶⁵), but are classified as euchromatin here.

§ The sum of lengths for finished sequence, estimated heterochromatic gaps, euchromatic gaps and unfinished clone gaps. The total length is only approximate because of uncertainty in gap sizes, particularly for heterochromatic gaps and centromeres.

|| Those in the tiling path but for which it has not been possible to obtain finished sequence. Unfinished sequence from these clones is deposited in public databases. These gaps are all listed at 50 kb, reflecting the approximate average size of the gap.

Box 2

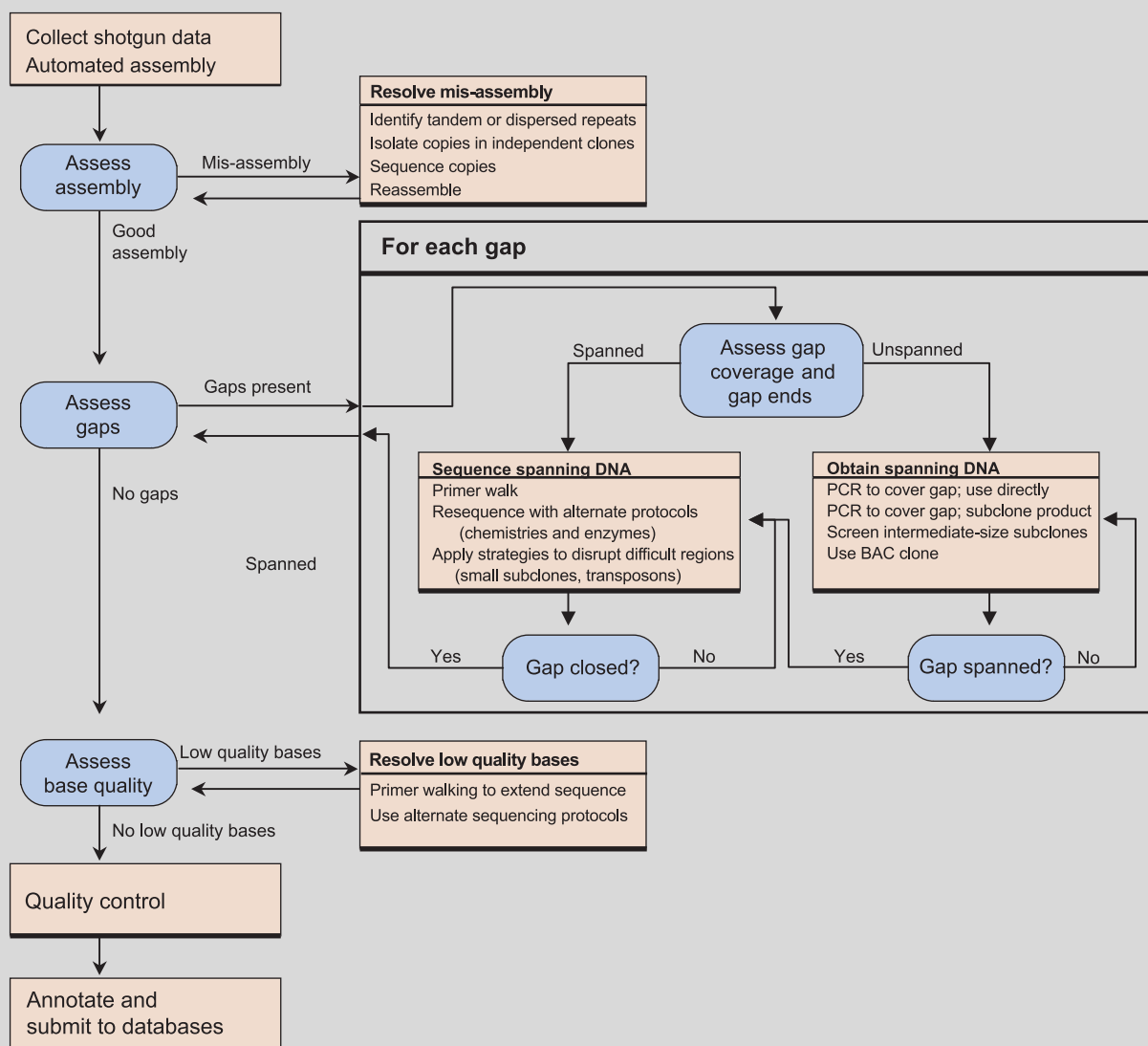
Finishing the sequence of clones

Sequencing of large-insert clones began with production of an initial assembly based on shotgun sequence data. For a typical BAC clone, we generated 6–10-fold coverage in paired end-sequences from random, small-insert (2–4 kb) plasmid clones and used computer programs (P. Green, unpublished and see ref. 68) to assemble the data into sequence contigs connected by linking information. Each base was assigned a quality score reflecting its predicted accuracy, based on the underlying shotgun sequence data. The assembly typically had gaps and low-quality regions, with the number varying greatly across clones. These regions are highly enriched in sequences that are difficult to clone or sequence and thus are not represented even after deep (6–10-fold) coverage with random reads. (These regions are also poorly covered by whole-genome shotgun strategies⁶⁹.)

The finishing phase converted this draft assembly into a high-quality continuous sequence by obtaining directed information. It involved iterative cycles of computational analysis and laboratory work. Box 2 Fig. 1 shows a simplified flowchart. The first step was to inspect the draft assembly for evidence of mis-assembly, arising from inappropriate merger of repeated sequences. Such evidence would include inconsistent patterns of linking among contigs, regions with unusually high coverage in sequence reads and bases with ‘high-quality

discrepancies’ among the underlying sequence reads. In general, sequence assembly is more straightforward for the clone-based hierarchical shotgun strategy than for the whole-genome shotgun strategy, because the use of clones avoids problems arising from polymorphism and from different copies of repeated regions elsewhere in the genome. Most clones passed assembly inspection, but some failed due to the presence of very similar local dispersed, tandem or inverted repeats. Careful inspection could resolve the problem in some cases, but specific strategies had to be devised in other cases. One approach was to isolate distinct copies of the repeat in subclones of intermediate size (10-kb plasmids or fosmids) and sequence these subclones. Box 2 Fig. 2 illustrates an initially mis-assembled BAC clone from chromosome Y that could be assembled correctly with careful editing.

The second step was gap closure. Because gaps tended to be enriched for problematic sequences, gap closure was challenging; it often required multiple attempts using a variety of alternative methods. Gaps were classified into two types: ‘spanned’ and ‘unspanned’. Spanned gaps were those for which the two flanking contig ends were linked by an end-sequenced plasmid. Most such gaps could be closed by primer-directed sequencing of the plasmid, serially extending the

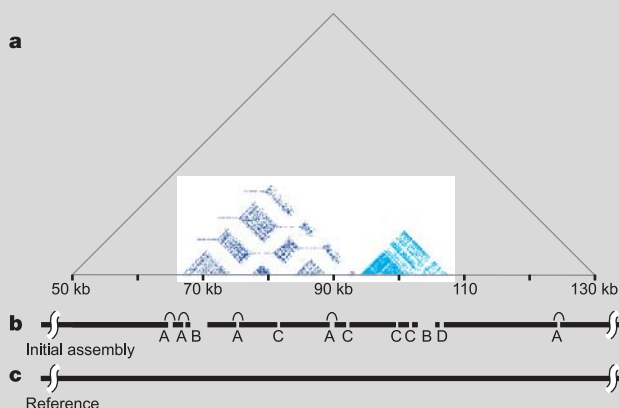


Box 2 Figure 1 Simplified flowchart for finishing of clones.

contig sequence into the gap. Sequence in the gap was often recalcitrant to the standard sequencing protocol (accounting for its absence from the initial shotgun data), making it necessary to use many alternative protocols (different buffers, enzymes and temperature conditions). Some gaps could not be closed by primer walking, because no suitable primer could be found (due to repetitive sequence near the end of the contig) or because sequencing chemistries were unable to penetrate certain secondary structures (such as some inverted repeats). Specialized strategies were used to obtain the missing sequence. For example, problems arising from secondary structure might be overcome by sequencing a small insert library⁷⁰ of random subclones with tiny inserts (~100–300 bp, referred to as a “shatter library”) or by sequencing from transposon insertions in the plasmid. Unspanned gaps arose where a contig end was not linked to any other contig. It was then necessary to infer adjacency and extend the sequence by other means. Techniques included PCR to other contigs, analysis of various types of subclones from the BAC, and primer walking directly on the BAC⁷¹. This battery of techniques succeeded in virtually all cases. (In 728 cases, there remains a small region of bases that could not be reliably sequenced; almost all fall in tandem repeat sequences and typically affect tens to hundreds of bases. These cases are annotated in the accessioned clones.)

The third step (which proceeded in parallel to gap closure) was the resolution of low-quality regions. This was accomplished by obtaining additional sequence reads from resequencing of existing shotgun subclones or from primer-directed sequencing.

The final step involved quality control. To confirm the accuracy of the overall assembly, the restriction digestion pattern of the BAC predicted from the finished sequence was compared with the pattern observed experimentally. To confirm accuracy at the nucleotide level, the finished sequence and supporting data were reviewed by human inspection and computational analysis. The finished sequence was then annotated and deposited in public databases.



Box 2 Figure 2 Illustration of a particularly challenging clone. **a**, Central portion of clone RP11-48811 is illustrated by a triangular dot plot. The base of the triangle represents 80 kb of a 152-kb insert. Each dot represents a perfect match of 20 bases. The region between 65 kb and 94 kb contains four copies of a directly repeated sequence of about 3 kb (horizontal lines), separated by imperfect short tandemly repeated sequence (diamond blocks of dots). The region between 94 kb and 107 kb contains tandemly repeated imperfect copies of a five-base sequence, unrelated to the previous sequence. **b**, Initial assembly of region after completion of shotgun data collection. Two mis-assemblies resulting from the long direct repeats and the absence of all copies (not shown) were resolved by manual editing, after which 12 gaps remained in the clone. Five of these (labelled A) were spanned by plasmid subclones and were closed by primer walking. Two gaps (labelled B) were larger; after initial walks failed, these gaps were closed by sequencing short insert libraries prepared from PCR products. Four other gaps (labelled C) were not spanned by plasmid clones but were closed by primer walks on PCR products. One gap (labelled D) was closed by primer walks and extensive manual editing. **c**, The finished clone with all gaps closed.

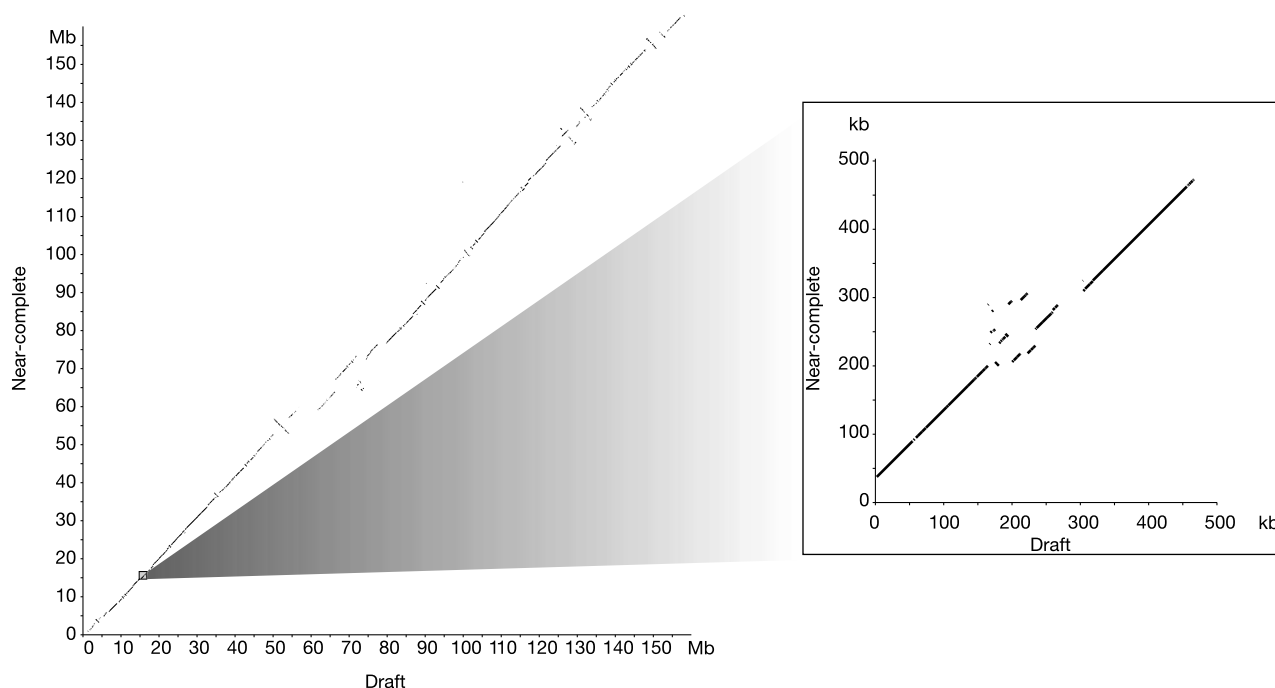


Figure 1 Comparison of previous draft sequence with current near-complete sequence of chromosome 7 (ref. 24). At large scale, there was good collinearity between draft and near-complete sequence, although some inversions were present in the draft due to lack of sufficient anchors in some regions. At finer scale, the draft sequence contained some

sequence contigs for which order and orientation were not known. The inset shows a region of 500 kb with sequence derived from three overlapping BACs. BACs at each end were finished at the time of draft assembly, whereas the middle BAC was at an early stage of shotgun coverage in which contigs were not yet ordered and oriented.

permanent foundation for biomedical research, it was important to assess its quality and to characterize its remaining defects. For this purpose, we used a number of comparisons and consistency checks.

Assessment of accuracy

Tests of accuracy were designed to detect potential problems that may have occurred in clone-based sequencing. This may include errors in assembling the finished sequence within individual clones, and errors in concatenating adjacent finished clones to create the final product. The analysis was complicated by the presence of polymorphism in the human population, because differences between sequence clones may reflect either errors or polymorphism.

Independent quality assessment. Quality assessment (QA) exercises were performed regularly throughout the HGP³¹. In the final stages, an independent group examined a random sample of finished clones by generating additional data and generating new assemblies³². Briefly, this QA analysis examined ~34 Mb and found an error rate of 1.1 per 100 kb for small events (≤ 50 bp, with average

size of 1.3 bp) and 0.03 per 100 kb for large events (> 50 bp). The small events consisted largely of single-base substitutions, whereas the remaining small and large events primarily concerned the number of consecutive copies of a tandem repeat³².

Analysis of clone overlap. We extended the QA analysis to a larger region (~174 Mb), by examining overlapping sequence between consecutive finished large-insert clones. If two such clones derive from the same copy of the human genome, any sequence differences in the overlap must reflect an error in one of the two clones. By comparing independent clones, this quality assessment method also has the ability to detect cloning artefacts. We examined 4,356 substantially overlapping clones derived from the same library; half are expected to be derived from the same haplotype and half from a different haplotype. We counted the number of single-base mismatches (ignoring insertion/deletions (indels)) in the overlapping regions. The resulting distribution (Fig. 2a) is bimodal. The first peak is consistent with expectation for clones from the same haplotype, with a sequencing error rate of $\sim 10^{-5}$ per bp. The second peak is consistent with the expectation for clones from different haplotypes, with a polymorphism rate of $\sim 10^{-3}$ per bp; this peak matches the distribution seen for clones from different libraries.

We then examined overlapping clones likely to be from the same haplotype (with no single-base mismatches) and counted the discrepancy rate for indels (Fig. 2b). The error rate (estimated as half the discrepancy rate) is ~ 0.55 events per 100 kb, with the vast majority being in tandem repeats. By contrast, clones from different libraries show a discrepancy rate that is at least 20-fold higher. Overall, the analysis indicates that the overall error rate (reflecting both sequence error and cloning artefacts) is 20–100-fold lower than the human polymorphism rate.

Analysis of junctions. We assessed longer-range integrity of the genome sequence by studying read pairs from large insert clones. Specifically, we created a fosmid library carrying randomly sheared human DNA and sequenced both ends of the insert of ~750,000 clones. Fosmid clones are particularly useful because their insert sizes cluster tightly around 40 kb, due to packaging constraints. We aligned the fosmid end sequences to the genome sequence. Both ends could be mapped to unique locations in the human genome in most cases (86%), and these two locations were within 39.5 ± 7.5 kb in 99% of cases. Some fosmids could not be uniquely placed because one or both ends consisted almost entirely of repeat sequence. Using the uniquely placed fosmids (which provide about eightfold clone coverage of the euchromatic genome), we sought to obtain independent confirmation of the order, orientation and adjacency of the junction between consecutive finished large-insert clones used to construct the genome sequence. The junction was considered 'supported' if spanned by one or more consistently placed fosmids. In all, ~97% of junctions were supported. About half of the remaining junctions were supported by fosmids with unique placement at one end but multiple placements at the other end. Overall, the analysis provided strong support for accuracy of the junctions underlying the current genome sequence.

Search for deletions. We next scanned the genome sequence for evidence of deletions of several kilobases in size, using the same fosmid data set. At each point, we calculated the 'apparent size' of each fosmid spanning the point (defined as the distance between the location of the end sequences in the current genome sequence) and then calculated the 'average apparent size' for all the fosmids spanning the point. We searched for regions where the observed size fell far below expectation (< 3.5 standard deviations (s.d.)), suggesting a large difference between the genome sequence and the source DNA for the fosmid library (Fig. 3). Such differences could reflect either an error in the genome sequence, a deletion in the fosmid clone, or a deletion polymorphism between the DNA sources. (Given the number of fosmids used, this analysis has ~50% sensitivity to detect deletions of 3–30 kb. Because the

Table 3 Chromosome arm length and contiguity in draft and reference sequence

Chromosome	Euch. length* (bp)	N50† draft§ (bp)	Build 35 N50 ref (bp)	N-average ref§ (bp)
1p	121,147,476	81,895	16,783,271	33,566,574
1q	104,135,370	45,843	56,331,646	36,675,159
2p	91,748,045	68,853	68,373,980	53,478,029
2q	148,270,183	50,481	84,213,156	54,482,973
3p	90,587,544	39,322	66,080,833	54,853,737
3q	106,018,194	35,734	100,530,261	96,935,077
4p	49,501,045	36,494	9,040,907	13,797,821
4q	138,910,172	31,876	92,070,735	66,386,026
5p	46,441,398	59,470	46,378,398	46,378,398
5q	131,416,467	81,416	41,199,371	33,564,217
6p	58,938,125	251,648	48,945,890	42,200,138
6q	109,037,573	150,424	61,695,806	46,408,435
7p	57,864,988	399,235	47,497,097	40,050,874
7q	97,763,150	298,612	64,426,257	46,810,648
8p	43,958,052	40,151	9,464,880	9,872,060
8q	99,316,773	37,528	57,155,273	47,945,192
9p	46,035,928	87,767	39,435,726	34,619,306
9q	74,393,339	43,983	40,394,264	29,078,785
10p	39,244,941	48,121	20,794,160	15,791,760
10q	93,788,686	47,401	30,112,613	31,833,318
11p	51,450,781	34,383	49,571,094	48,044,101
11q	80,001,602	42,527	17,911,127	26,070,918
12p	34,747,961	197,985	27,615,668	23,435,010
12q	96,306,849	47,272	32,815,934	29,605,325
13p	acro arm	n/a	n/a	n/a
13q	96,274,979	70,497	67,740,325	54,830,719
14p	acro arm	n/a	n/a	n/a
14q	88,298,584	1,370,997	88,290,585	88,290,585
15p	acro arm	n/a	n/a	n/a
15q	82,078,915	30,303	53,619,965	38,049,097
16p	35,143,302	160,390	25,336,229	20,462,803
16q	43,883,952	86,933	42,003,582	40,305,188
17p	22,187,133	114,901	21,163,833	20,341,190
17q	56,487,608	82,866	11,472,733	15,591,618
18p	15,400,898	59,951	15,400,898	15,400,898
18q	59,352,257	50,087	33,548,238	26,073,241
19p	26,923,622	82,369	15,825,424	12,506,733
19q	33,888,028	167,408	31,383,029	31,383,029
20p	26,267,569	1,436,102	26,259,569	26,259,569
20q	34,402,734	1,301,134	26,144,333	21,428,992
21p¶	490,223	n/a	490,223	490,223
21q	33,684,323	28,515,322	28,617,429	24,743,931
22p	acro arm	n/a	n/a	n/a
22q	35,224,709	23,048,103	23,276,302	16,327,958
Xp	58,465,033	173,718	33,063,353	22,383,515
Xq	93,359,231	277,548	27,718,692	25,766,623
Yp	11,237,315	5,778,849	6,265,435	4,331,076
Yq	15,464,376	1,026,317	10,002,238	8,061,778
All arms	2,879,539,433	82,663	38,509,590	40,970,092

*Chromosome arm lengths refer to estimated length of euchromatic portions of each arm.

†N50 denotes the contig length x (for a chromosome arm or entire genome) such that half of all nucleotides reside in contigs of length at least x .

‡'N50 draft' reports this number for the draft sequence¹⁵.

§The value for the near-complete reference sequence reported here.

||Average contig length in the near-complete sequence for a randomly chosen nucleotide (or, equivalently, average length contigs weighted by length).

¶Chromosome 21p is an exception to the generalization that the acrocentric arms only contain heterochromatin—there is a 281-kb contig within chr 21p11.2.

methodology cannot detect deletions larger than a fosmid, we also analysed discrepant fosmid links, which could reflect deletions. See Methods in Supplementary Information.)

We found 242 candidate regions, with suggestive evidence for deletions (average apparent size ~ 5 kb). These regions were then scrutinized by alignment with the recently obtained draft sequence of the chimpanzee genome (R. H. Waterston, personal communication). Because the human and chimp genomes align with relatively few large indels (indels >2 kb occur at ~ 1 per 100 kb), this comparison should highlight true deletions. The chimpanzee comparison supported the presence of deletions in 35% of cases. A subset of these was then tested by polymerase chain reaction (PCR) analysis of genomic DNA from multiple individuals. Roughly two-thirds appear to represent polymorphic deletions in the human population and one-third represent actual errors in the current genome sequence. Overall, the results indicate that the current

genome sequence is likely to contain perhaps 50–100 erroneous deletions (average size ~ 5 kb), which could be due to assembly errors or mutations occurring during propagation of large insert clones. Analysis of a larger collection of fosmids could probably pinpoint the majority of these errors, allowing them to be corrected.

Assessment of coverage

Tests of coverage were designed to measure the proportion of the euchromatic genome missing from the current genome sequence, by assessing the presence of independently sampled human sequences such as complementary DNA clones and random genomic clones.

Analysis of cDNAs. We tested for the presence of known cDNA sequences from public databases (REFSEQ³³ and MGC³⁴). The analysis³⁵ involved 17,458 distinct gene loci spanning 925 Mb of genomic sequence. The vast majority (99.74%) could be confidently aligned to the current genome sequence over virtually their complete length with high sequence identity (at a level consistent with the expected polymorphism rate and the performance of the alignment program). A few of these (0.5%) showed strong alignment to more than one locus. A few others (0.04%) showed unusually high sequence difference ($>2\%$), but these were nearly all immunologically related genes (such as major histocompatibility loci and immunoglobulin-related loci) known to be highly polymorphic.

We examined the remaining cases (0.28%). The cDNA sequence appeared to be completely absent in 0.06% of cases and partially absent, with a contiguous segment missing, in 0.23% of cases. For

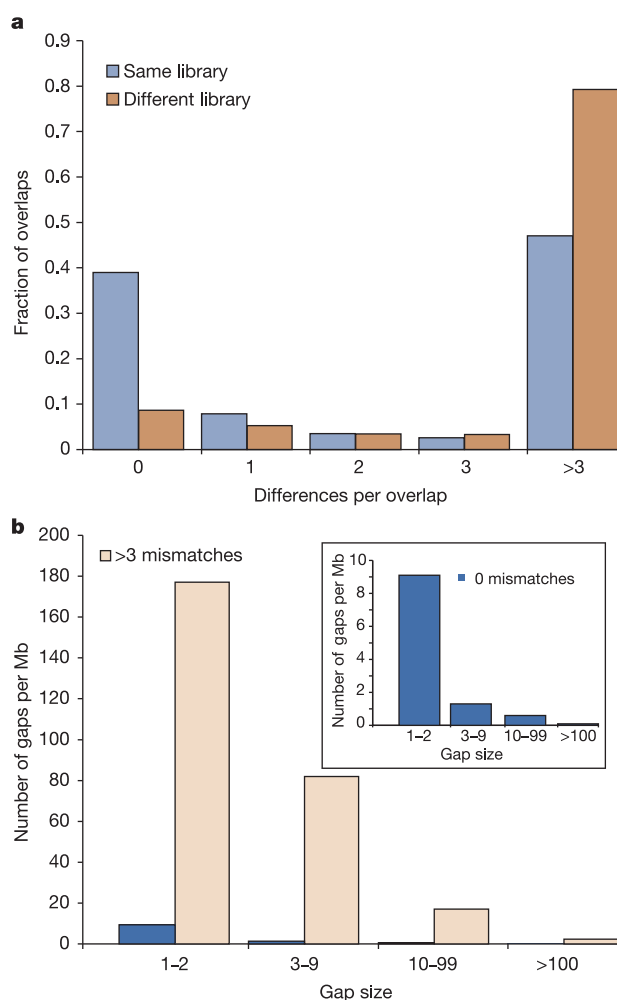


Figure 2 Assessment of potential errors by analysis of BAC overlaps. **a**, Single-base differences between overlapping finished BAC clones (with ≥ 5 kb overlap). The number of single-base differences in overlaps for clones from the same library and from different libraries is plotted. The results are consistent with half of the clones from the same library representing identical underlying DNA sequence with low error rate, and half representing different haplotypes as expected. **b**, Insertion/deletion (indel) differences between overlapping clones. The number of indels per Mb for a given size range is compared for clones with no single-base mismatches (presumed to be derived from the same haploid source) and >3 single-base mismatches (presumed to be derived from different haploid sources). Indels in the former class primarily represent errors in finished sequence; they occur at ~ 20 -fold lower frequency (inset) than indels in the latter class, which primarily represent polymorphic differences.

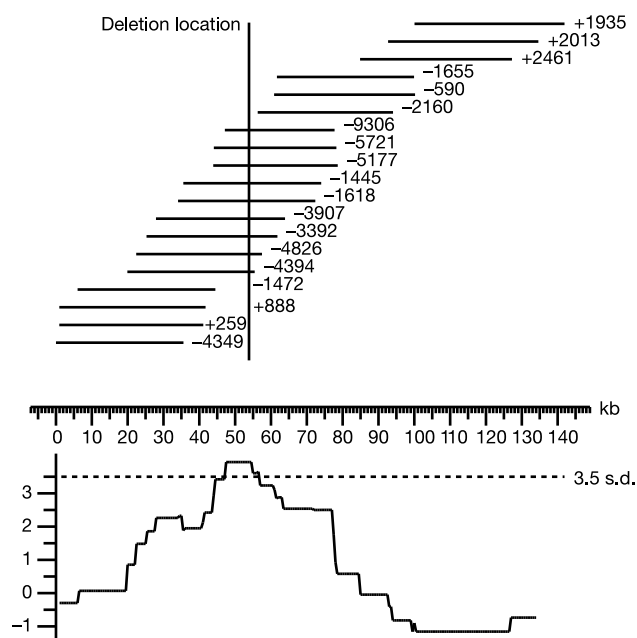


Figure 3 Detection of potential insertions or deletions using paired-end fosmid reads. The top portion shows fosmids along a region of chromosome 10 (centred at nucleotide 46,915,451), mapped by virtue of their paired-end sequences. The difference between inferred length, calculated from the location of fosmid ends in finished sequence, and average length for the entire library, is shown to the right of each clone. For each point, the standard deviation of the local average difference for all spanning fosmids is plotted below; the threshold of 3.5 standard deviations is indicated by a dotted line. The region from 45 to 55 kb is inferred to contain a length difference between the fosmids and finished sequence. Comparison with available chimpanzee sequence further localized the difference (vertical line). Experimental analysis (PCR from clones used for finished sequence and the fosmid library, as well as from 24 random humans) confirmed the difference, and showed that it is due to an insertion/deletion polymorphism of 5.8 kb. The majority of length differences detected by this analysis appear to represent polymorphisms, not sequence errors.

almost all of completely absent cDNAs, the genomic location of the gene was known or could be inferred and corresponds to a gap in the current genome sequence. For the partially absent cDNAs, more than half of the cases lie adjacent to gaps. The remainder may represent either errors in the current genome sequence or polymorphic deletions; these are being investigated further. Overall, the proportion of cDNA sequence that is missing from the genome sequence is only 0.08% of the total. This may underestimate the proportion of genome missing from the finished sequence, however, because focused efforts were made to capture genomic sequence containing missing messenger RNAs.

Analysis of random genomic plasmids. As an additional and broader test of coverage, we analysed paired end-sequences from 5,000 small-insert (3–4 kb) plasmids generated as part of a human single nucleotide polymorphism (SNP) discovery project (see

Methods). After excluding heterochromatic repeats and other artefacts, we found that 99.3% of the reads could be reliably aligned to the finished sequence. For 0.6% of the reads, neither end could be aligned; these probably lie in known gaps. For another 0.1% of the reads, exactly one end could be placed; some fell next to known gaps, whereas others appear to represent indel differences between the reference sequence and the source DNA for the plasmid library. The overall analysis indicates that <1% of the euchromatic genome is missing from the finished sequence. Together, the cDNA and plasmid analyses indicate that the current genome sequence contains more than 99% of the euchromatic portion of the human genome.

Characterization of remaining gaps

The current genome sequence contains 341 gaps, which could not

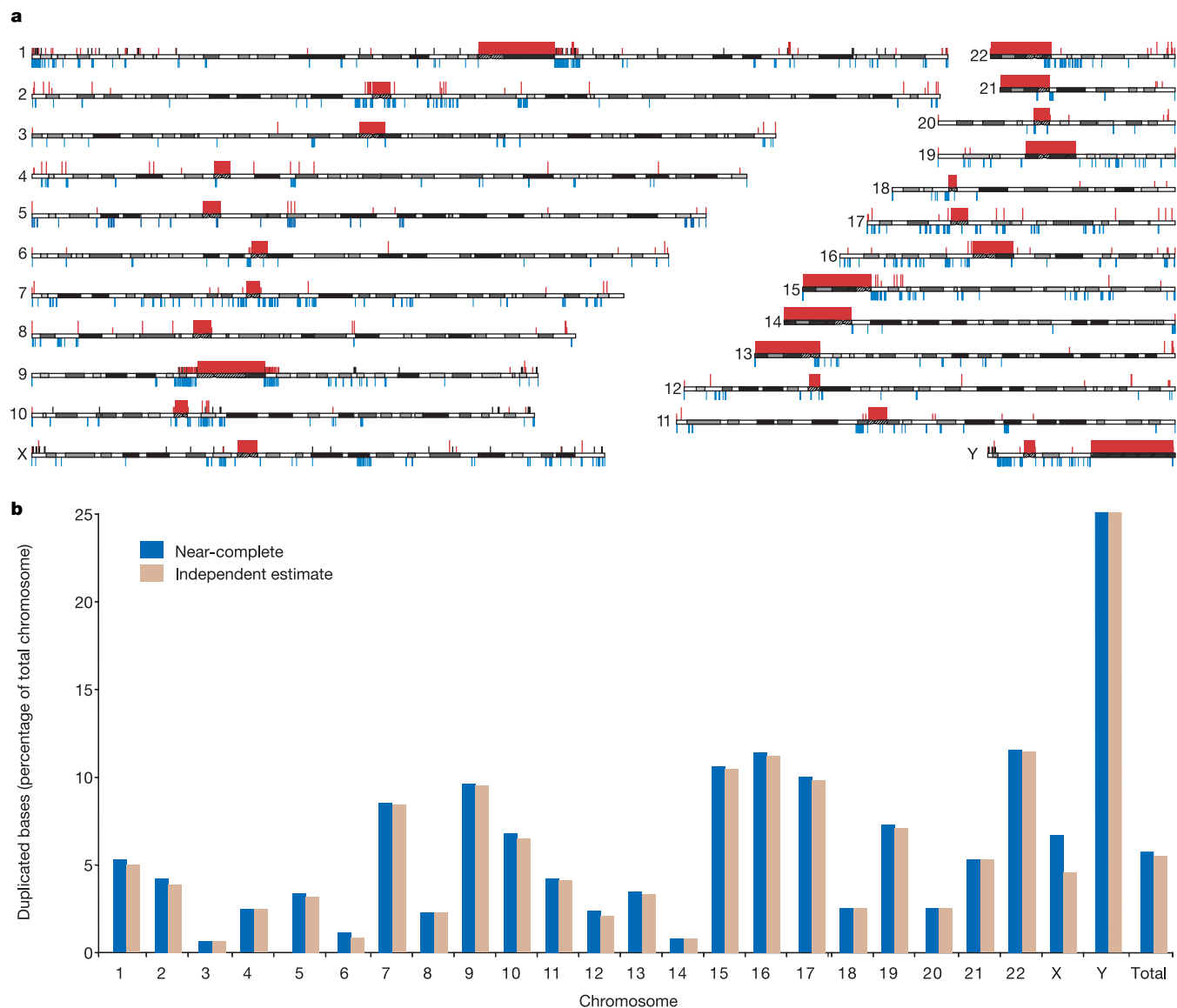


Figure 4 Segmental duplications across the genome. **a**, Segmental duplications and sequence gaps across the genome. Segmental duplications are indicated below the chromosomes in blue (length ≥ 10 kb and sequence identity $\geq 95\%$). Large duplications are shown to approximate scale; smaller ones are indicated as ticks. Sequence gaps are indicated above the chromosomes in red. Large gaps (>300 kb) are shown to approximate scale; smaller gaps are indicated as ticks with those that are 50 kb or smaller shown as shorter ticks. Unfinished clones are indicated as black ticks. **b**, Percentage of large segmental duplications by chromosome. This count includes both interchromosomal

and intrachromosomal duplications with length ≥ 1 kb and sequence identity $\geq 90\%$. The blue bars show the result of direct analysis of near-complete sequence. The gold bars show an independent estimate⁶⁵ using whole-genome shotgun data to correct for potential mis-assembly of such segmental duplications. The strong agreement suggests that most segmental duplications are properly represented in near-complete genome sequence. The discrepancy for chromosome X is probably a result of errors in the independent estimate, due to limited coverage and diversity of data from this chromosome¹⁵.

be closed with available techniques. We briefly describe the nature of these gaps and discuss the prospects for eventual closure. (See Supplementary Information Notes 2 and 4.)

Heterochromatic regions (33 gaps). The heterochromatic regions of the human genome were not targeted by the HGP, because their

highly repetitive properties make them largely refractory to current cloning and sequencing strategies. There are 33 heterochromatic regions falling into four types. The 24 centromeres (~50 Mb) consist largely of alpha satellite repeats, of which ~15 types exist; these monomeric repeats are arranged into higher-order arrays

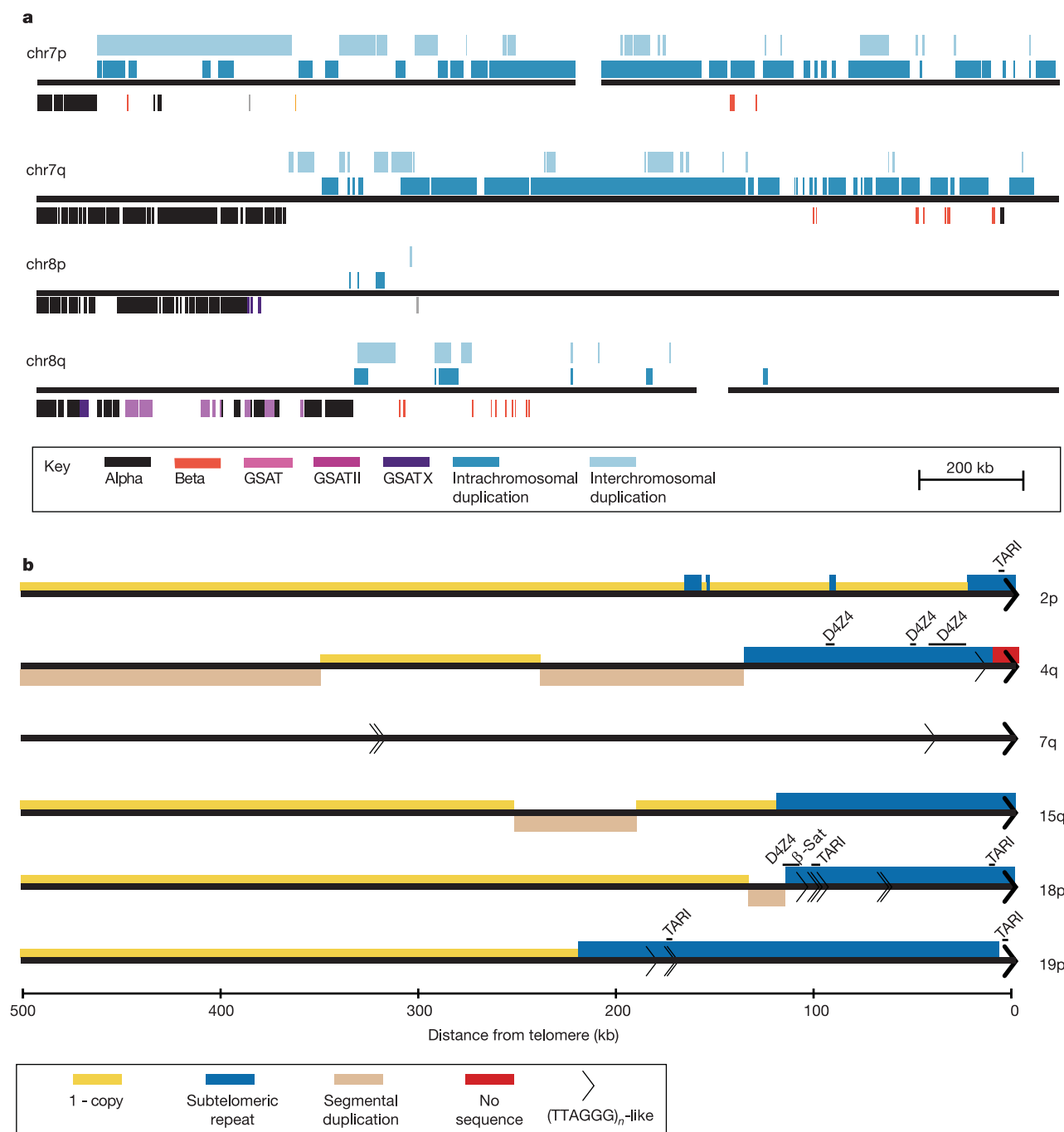


Figure 5 Examples of repeat structure near centromeres and telomeres. **a**, Repeats in pericentric regions of chromosomes 7 and 8. The most proximal regions are crowded with alpha satellite sequences and other centromeric repeats; composition, density and order may vary considerably between chromosome arms⁶². Just outside this region, there is usually a high density of inter- and intra-chromosomal duplication. For details, see text and refs 39, 40, 66 and 67. **b**, Sequence organization in human subtelomeric DNA regions. The terminal repeat tract consists of 2–15 kb of simple repeat sequence (TTAGGG)_n and is indicated by the black arrow at right. Short (50–250 bp) and often degenerate (TTAGGG) tracts (internal black arrows) are highly enriched (>25-fold) in subtelomeric DNA relative to elsewhere in the genome. A subtelomeric repeat (Srpt)

region (blue) consists of a mosaic patchwork of segmentally duplicated DNA tracts that occur in two or more subtelomeric regions and range in size from <10 kb to >300 kb. TAR1, D4Z4 and beta satellite sequences are frequently associated with Srpt regions. Proximal to the Srpt region is chromosome-specific genomic DNA, typically with a high GC content and high gene density. Stretches of segmentally duplicated DNA that occur only once within subtelomeric regions (tan) are interspersed with 1-copy subtelomeric DNA (yellow) in a telomere-specific fashion. Overall, segmentally duplicated DNA comprises approximately 25% of the most telomeric 500 kb of the chromosome, a fivefold enrichment over the genome-wide average.

distinct to specific chromosomes, which are tandemly repeated with slight sequence variations. The three secondary constrictions are immediately adjacent to the centromere on chromosome arms 1q, 9q and 16q and contain various satellite repeats (beta, gamma, satellite I, II, III). The five acrocentric chromosome arms 13p, 14p, 15p, 21p and 22p encode the 5S, 18S and 28S ribosomal RNA genes, which lie on a 43-kb sequence present in ~50 tandem copies on each arm and are flanked by additional repeats arranged in complex structures. Finally, there is a single large region on distal Yq composed primarily of thousands of copies of several repeat families. The heterochromatic regions all tend to be highly polymorphic in length in the human population.

Euchromatic boundary regions (35 gaps). The euchromatic regions of the human genome are bounded proximally by heterochromatin and distally by a telomere consisting of several kilobases of the hexamer repeat TTAGGG. We examined the current genome sequence for evidence of the expected boundaries on the 43 euchromatic arms. (See Supplementary Information Note 4.) At the proximal ends, 30 of the 43 cases show sequence characteristic of either heterochromatin or immediately flanking regions (such as higher-order centromeric repeats, stretches of at least 10 kb of monomeric alpha satellite repeat or other pericentromeric repeats). We cannot exclude the possibility that there is additional unique

sequence between this point and the proximal heterochromatin; but efforts to extend the finished sequence further were unsuccessful. In the remaining 13 cases, the finished sequence contains no evidence of heterochromatin-related sequence. At the telomeric ends, 21 of the 43 cases show continuous sequence extending to the telomeric repeat. This sequence was typically obtained by isolation and sequencing of half-YAC clones spanning to the telomere³⁶. An additional 18 cases are sequence gaps, in which half-YACs reaching to the telomere were isolated but finished sequence could not be obtained. The remaining four cases are physical gaps, in which large-insert clones extending to the telomere could not be obtained.

Euchromatic interior regions (273 gaps). The remaining gaps are located within the current genome sequence. These consist of 215 physical gaps for which no clones could be isolated, and 58 sequence gaps for which clones were found but reliable finished sequence could not be obtained. The physical gaps are greatly enriched in regions of segmental duplication (Fig. 4a). Roughly half of these gaps (52%) are flanked by segmental duplications with >90% sequence identity, although such duplications comprise only ~5.3% of the euchromatic genome (Fig. 4b). Such segmental duplications are especially frequent in pericentromeric regions, and gaps are notably more frequent in these regions. The association of gaps with segmental duplications is examined in detail elsewhere³⁷.

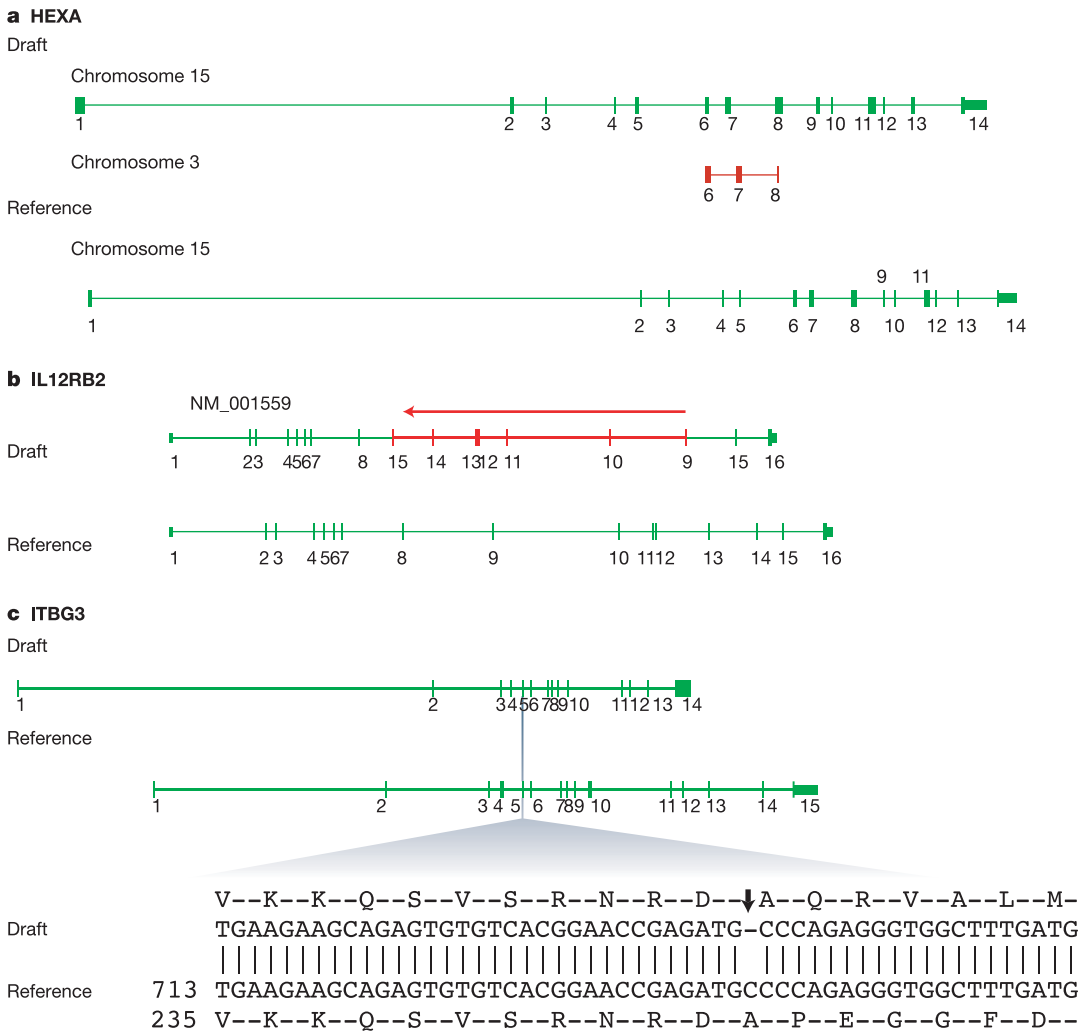


Figure 6 Examples of genes corrected in the near-complete sequence. **a**, In draft sequence, exons 6, 7 and 8 of hexosaminidase A (HEXA) gene were present on chromosome 3 in addition to their correct location on chromosome 15. **b**, In draft sequence, interleukin 12 beta-2 receptor (IL12RB2) contained an inversion of several

internal exons and a duplication of exon 15. **c**, In draft sequence, a single-base deletion in exon 5 of integrin beta-3 precursor (ITBG3) caused a frameshift at amino acid 247. The terminal exon was also missing.

The most extreme case occurs near the centromere of chromosome 9. The most proximal 5 Mb on 9p and 4 Mb on 9q comprise a mere 0.3% of the genome, but account for ~12% of the physical gaps in the euchromatic sequence. These two pericentric regions are unique in the genome with respect to density of segmental duplication and the average degree of intrachromosomal sequence identity (98.7%), and the two regions have many highly similar sequences in common. The high sequence similarity between the two regions is likely to be the reason for a polymorphic inversion of the centric heterochromatin on chromosome 9, present at a frequency of ~1% in the human population²⁸. Other proximal regions also show a higher-than-average density of gaps. For example, the proximal 2 Mb on the remaining 41 euchromatic arms comprise 2.9% of the genome but harbour 13.3% of the gaps. Nearly all of these proximal gaps are flanked by segmental duplications (Fig. 5a). There is also a clustering of such gaps in subtelomeric regions. The terminal 1 Mb on the 43 euchromatic arms represents 1.5% of the genome, but contains ~14% of the total gaps; nearly all of these gaps are also flanked by segmental duplications (Fig. 5b).

Closing the remaining gaps. Although the euchromatic genome sequence has reached a much higher degree of completion than had been anticipated, it still remains incomplete with ~1% of the euchromatin residing in 308 gaps. These represent regions that could not be reliably mapped, cloned and sequenced with current methods. Rather than applying further brute force, it is now time to develop focused strategies to resolve the regions.

The remaining euchromatic gaps probably reflect two major issues. The first pertains to regions harbouring segmentally duplicated sequence. Such regions are challenging to map because it can

be extremely difficult to discern whether two clones with small sequence differences represent different loci or different alleles at a single locus. This challenge was eventually resolved for chromosome Y (ref. 23) (which is especially rich in segmental duplication) by exploiting the fact that the chromosome is haploid in males. By using DNA from a single haploid source, it was possible to rely on differences at only a handful of nucleotides to distinguish repeated sequences. This approach could be applied to the rest of the genome by using appropriate haploid sources, such as a hydatidiform mole or monochromosomal hybrids. (In both instances, use of parental controls to guard against being misled by somatic rearrangements would be well advised.) It may be useful to test these approaches on individual chromosomes. The second issue is that some gaps are likely to correspond to regions that cannot be efficiently propagated in current large-insert vectors and hosts. It may be useful to test new kinds of large-insert libraries for clones containing unique sequences not contained in the current human genome sequence (perhaps seeded by probes derived from random small-insert genomic plasmids, as discussed above). In addition, genome completion may benefit from long-range mapping techniques such as optical mapping³⁸, which may provide independent information about difficult regions.

Completing the euchromatic sequence is an important goal, but is clearly now a research effort rather than a high-throughput project. Sequencing the human heterochromatin poses an even greater challenge. The current human sequence penetrates only the periphery of the heterochromatin—for example, the pericentric regions on a few chromosome arms^{39,40}. This progress has required concerted efforts with specialized mapping techniques and painstaking assembly. The fundamental issue is that current shotgun

Table 4 Human paralogous genes

Chromosome	Cluster size in human genome	Minimum size in ancestral genome	Genes involved in recent duplications	Gene family
2	30	8	26	Immunoglobulin K chain V
11	64	50	23	Olfactory receptor
19	23	5	19	KRAB zinc-finger protein
14	21	9	15	Immunoglobulin heavy chain
9	16	4	13	Interferon α
1	34	25	13	Olfactory receptor
19	18	9	13	Leukocyte and NK cell immunoglobulin-like receptors
22	20	11	12	Immunoglobulin λ chain V-region
1	13	3	11	PRAME/MAPE family (cancer/germ line antigen)
16	11	2	11	Immunoglobulin heavy chain
19	10	1	10	Pregnancy-specific β -1-glycoprotein
11	59	54	10	Olfactory receptor
X	9	1	9	Hypothetical gene LOC255313 expressed in testis
17	13	9	8	Olfactory receptor
4	9	3	7	UDP-glucuronosyltransferase; steroid metabolism
12	14	8	7	Taste receptor, type 2
19	16	10	7	FDZF2-like KRAB zinc-finger protein
X	7	1	7	SSX-like KRAB zinc-finger protein (CT antigens)
X	8	2	7	MAGE (CT antigens)
Y	7	1	7	Testis-specific Y-encoded (TSPY) protein
4	9	5	7	CXCL1/MIP2-like small chemokines
7	6	1	6	Postmeiotic segregation increased-2 (DNA mismatch repair)
8	6	1	6	FLJ00326 hypothetical protein
11	6	1	6	TRIM48, testis-specific RING finger protein
16	6	1	6	Tumour protein p53-inducible gene, TP53TG3
Y	6	1	6	Testis-specific Y-encoded (TSPY) protein
6	7	3	6	Butyrophilin subfamilies 2 and 3
11	23	19	6	Olfactory receptor
19	18	14	6	Gonadotropin-inducible transcription repressor-2-like
7	5	1	5	Williams Beuren syndrome chromosome region 19 protein
8	5	1	5	Exonuclease GOR
11	5	1	5	Yeast Ssu72p-like protein
14	6	2	5	Immunoglobulin α , δ , γ chains
16	5	1	5	Metallothionein 2A-like
17	5	1	5	Growth hormone gene cluster
19	5	1	5	Testis-specific transcriptional repressor
19	5	1	5	Choriogonadotropin beta (placental hormones)
X	6	2	5	GAGE (CT antigens)
X	5	1	5	XAGE (CT antigens)
X	5	1	5	Sarcoma antigen SAGE (CT antigens)
X	5	1	5	SPAN-X; sperm protein (CT antigens)

The above table includes clusters of human paralogous genes having at least five genes involved in recent gene duplications. Recent is defined as divergence $K_S \leq 0.3$ (see text), corresponding roughly to the divergence of primate and rodent lineage. Ancestral cluster size is the minimum required to account for existing clusters in human, assuming no gene losses in the human-specific lineage.

strategies are poorly suited to assembling large, highly repetitive regions. The hierarchical shotgun strategy faces the challenge of accurate assembly of individual BACs and accurate overlap of BAC clones, with the underlying data consisting of nearly identical sequence; the whole-genome shotgun strategy compounds these problems. Conceivably, the hierarchical strategy could be adapted as was done for repetitive regions of chromosome Y. Approaches might include the use of the following: haploid DNA sources to restrict the problem to a single haplotype; single chromosome sources to avoid confusion among related centromeres on different chromosomes; sheared BAC libraries to avoid biases caused by the unusual distribution of restriction sites within the repeat sequences; assembly based on rare base differences that distinguish near-identical repeats; cloning vectors that minimize rearrangements; and subclone libraries of varying insert lengths. Such an approach will also require ensuring accurate recovery and stability of heterochromatic regions in large-insert clones. Even so, the path is likely to be arduous and expensive to obtain regions of uncertain information content. Alternatively, it may be possible to develop new approaches. These might include methods to obtain much longer effective read lengths, directed reads from known locations and long-range mapping information about the location of rare base differences among repeat copies (such as optical mapping³⁸ or padlock probes⁴¹).

Examples of utility of near-complete sequence

The present genome sequence enables far more precise analyses of the human genome, especially those that depend sensitively on high accuracy and near-completeness. Rather than revisit all of the analyses in our initial analysis of the human genome, we have chosen four examples that illustrate the utility of the current near-complete sequence.

Segmental duplications

The human genome is notable for its high proportion of recent segmental duplications. They are of great medical interest because their unusual structure often predisposes them to deletion or rearrangement with consequent phenotypic effects; prominent examples include the Williams syndrome region (7q), Charcot-Marie-Tooth region (17p), DiGeorge syndrome region (22q) and the AZF-C region (Y)⁴². Some regions of segmental duplication have also recently been shown to be evolutionary nurseries in which coding sequences are undergoing strong positive selection⁴³. Accurate analysis of segmental duplications was previously impossible

because the draft sequence also contained a high degree of artefactual duplication. This difficulty was recognized at the time and the approximate proportion of true and artefactual duplication was inferred indirectly. With near-complete sequence, the artefacts are now largely eliminated and true segmental duplications can be reliably studied.

On the basis of the current sequence, segmental duplications cover ~5.3% of the euchromatic genome. (Here, segmental duplications are counted as regions that are not transposable element copies, are ≥1 kb in length and have sequence identity ≥90%; this corresponds to duplication within the past ~40 million years.) The proportion of segmental duplication and the degree of sequence identity are clearly substantially higher in the human genome than in the mouse⁴⁴ or rat⁴⁵ genomes (although precise figures for the rodent genomes must await finished sequence). The use of large insert clones, representing a single haplotype, was critical in resolving these regions. The distribution of segmental duplication varies widely across chromosomes, as does the proportion of intrachromosomal versus interchromosomal duplications¹⁵ (Fig 4b). The most extreme case is chromosome Y, which carries segmental duplication along >25% of its total length and includes blocks as large as ~1.45 Mb with sequence identity of ~99.97% (ref. 23). In addition, many pericentromeric and subtelomeric regions are rich in dispersed segmental duplications (Fig. 5), apparently resulting from a steady bombardment of insertional translocations⁴⁶. Although most regions of segmental duplication have now been sequenced, ~10% of them lie in the remaining gaps in the current sequence and will require further work to elucidate, as discussed above.

Protein-coding genes

A central goal of genome analysis is the comprehensive identification of all human genes. This task remains challenging, but is greatly aided by the near-complete sequence together with other improved resources (such as expanded cDNA collections, genome sequence from other organisms and better computational methods). The current version of the human gene catalogue (Ensembl 34d) contains 22,287 gene loci (with a total of 34,214 transcripts, corresponding to 1.54 transcripts per locus), consisting of 19,438 known genes and 2,188 predicted genes. These gene loci have a total of 231,667 exons, with ~10.4 exons per locus and ~9.1 exons per transcript. The total length covered by the coding exons is ~34 Mb or ~1.2% of the euchromatic genome; the untranslated regions of the transcripts are estimated to cover another ~21 Mb or ~0.7% of the euchromatic genome.

Comparison of the initial and current gene catalogues highlights the substantial improvement. Many of the earlier gene models were erroneous due to defects in the draft sequence. Examples resulting from a duplication, inversion and premature stop codon are shown in Fig. 6. The improvement can be quantified by mapping the current gene models onto the draft sequence, to determine whether they could have been accurately identified. Of the transcripts in the current gene catalogue, 58% have at least one error when mapped onto the draft sequence. For 39% of transcripts, there is at least one

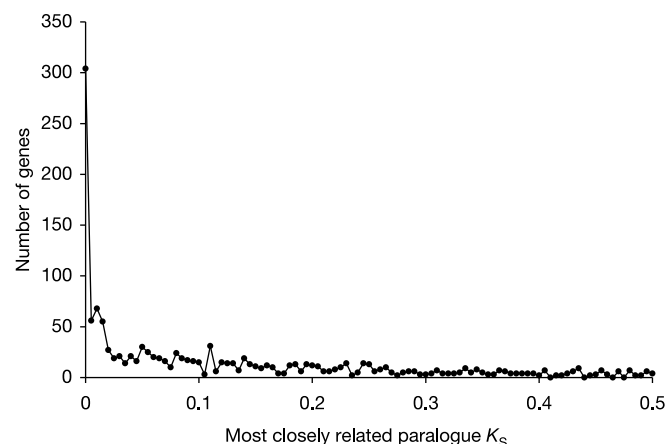


Figure 7 Distribution of K_S for recent gene duplications. K_S , the number of synonymous substitutions per synonymous site, was estimated for each gene from alignment with its most closely related human paralogue. This provides an indication of evolutionary time since divergence.

Table 5 Categories of recently arising human pseudogenes

	Inactivating mutations in the reference sequence		Total
	1 mutation	>1 mutation	
Chimp pseudogene	8	19	27
Human pseudogene	5	0	5
Human allelic null	1	0	1
Sequence error	1	0	1
Total	15	19	34

Potential pseudogenes were identified by searching orthologous segments in human, mouse and rat (see text). Frameshifts and nonsense mutations in putative human pseudogenes were experimentally investigated by resequencing the original human BAC clones (to look for sequence errors), as well as sequencing 24 unrelated individuals (to assess whether apparent mutation is a polymorphism) and chimpanzee (to assess timing of loss of function).

exon that is absent or incorrectly ordered due to defects in the draft. For the remaining 19% of transcripts, the exons are all present and correctly ordered, but there are one or more nucleotide errors.

Automated gene annotation has now been complemented by manual annotation of most chromosomes, based on a careful review of gene structure and examination of expressed-sequence-tag (EST) and transcript evidence. Such analysis has been completed for 18 chromosomes (2, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 17, 18, 19, 20, 21, 22, X and Y; refs 17, 20–30 and http://vega.sanger.ac.uk/Homo_sapiens/), with the remainder in the press or in preparation) comprising 1.7 Gb of the euchromatic genome. Although this annotation has further improved the quality of the gene models⁴⁷ (by dealing with special cases and unusual features not yet handled in the automated programs, and resolving instances of conflicting experimental data), it has not significantly affected the total gene count for these chromosomes.

On the basis of available evidence, our best estimate is that the total number of protein-coding genes is in the range 20,000–25,000. The lower bound seems secure, based on the number of currently known genes (19,599). The upper bound is based on estimates of the number of additional genes. Despite intense automated and manual analysis using cDNA, EST and cross-species homology, only 2,188 gene predictions have been added to the known set. This predicted set is likely to represent substantially fewer than 2,000 true genes, owing to fragmentation and false predictions arising from pseudogenes. For example, the predictions tend to have fewer exons per transcript than known genes (~4.7 versus ~9.7) and to encode shorter open reading frames (~847 versus ~1,487 amino acids). On the other hand, the set is likely to be incomplete because some protein-coding genes have surely continued to escape detection. The most problematic cases would be genes that have very short open reading frames (<100 amino acids), consist of single exons or evolve very rapidly. Even if we assume that such genes comprise 10% of the total (which seems a generous overestimate, given our current understanding of the human and other genomes^{48–50}), the total gene count would remain below 25,000. The range of 20,000–25,000 is also consistent with recent estimates (J. Weissenbach, unpublished) of the number of protein-coding genes based on cross-species homology (using the Exofish method⁵¹).

In our initial analysis of the draft sequence¹⁵, we estimated the count of human protein-coding genes at roughly 30,000. The estimate was derived as follows. We used computational analysis to generate an initial gene catalogue with ~32,000 entries, consisting of ~15,000 known genes and ~17,000 gene predictions. We estimated that the catalogue actually corresponded to ~24,500 actual genes, based on estimates of the rate of various types of errors such as fragmentation and false positive predictions (due largely to limitations of the draft sequence, such as imperfect recognition of pseudogenes and unknown order and orientation). We then adjusted the estimate to account for the proportion of genes estimated to be absent from the initial catalogue due to incomplete coverage of the genome and imperfect computational methods, resulting in a figure of ~31,000 genes.

With the current high-quality sequence, it is now possible to revisit this earlier analysis. We directly compared the previous gene models with the current gene models, to determine whether our previous estimates of the various error rates were correct. It is clear that the main reason for the earlier overestimate is that the fragmentation rate was substantially underestimated. The fragmentation rate is defined as the average number of the previous gene models that map to the same true gene; we assessed it by mapping to the current gene catalogue. The fragmentation rates for 'known' and 'predicted' genes were estimated in our earlier paper¹⁵ at ~1.0 and ~1.4, whereas our current analysis indicates that they should have been ~1.3 and ~1.7. This correction alone would bring our previous estimate to ~24,000. Small differences in the estimated

rate of false positive and negative predictions account for the remainder of the discrepancy.

It should be emphasized that the count above refers to the count of protein-coding genes. It does not include known non-coding RNAs, such as transfer RNAs, ribosomal RNAs, small nucleolar RNAs (snoRNAs) and microRNAs^{52–54}. In addition, there is evidence that the human genome gives rise to many additional RNA transcripts⁵⁵. It is unclear whether most such transcripts have specific biological functions or reflect reproducible transcriptional noise; few contain substantial open-reading frames and thus they are unlikely to encode proteins. There is a need for reliable experimental and computational methods for comprehensive identification of non-coding RNAs.

Finally, the near-complete sequence makes it possible to undertake systematic searches for pseudogenes. Automated annotation of chromosomes has focused primarily on identifying large pseudogenes of more recent origin. Recent published studies have used more sensitive methods to detect smaller and older pseudogenes and have already identified ~20,000 processed and unprocessed pseudogenes⁵⁶. This is surely still an underestimate, because such analysis will miss pseudogenes that are extremely old or that contain primarily untranslated regions. The total number of pseudogenes is thus likely to exceed the total number of functional genes. A particular type of pseudogene (recently arising non-processed pseudogenes) is discussed in more detail below.

Gene birth in the human lineage

The birth of new genes is of interest because it provides raw material for adaptive evolution, with extra copies of genes able to undergo functional divergence in response to positive selection. The quality and completeness of the current sequence make it possible to study this question; such analysis would have been unreliable with the earlier draft sequence, because the extensive artefactual local duplication would have given rise to many false positives.

We searched for clusters of nearby homologous genes, indicative of local gene duplication. The divergence between such genes was assessed at sites likely to be selectively neutral, by measuring the estimated substitution rate per synonymous site (K_S). We looked for nearby human gene pairs differing from one another by $K_S < 0.30$, implying that each differs from the common ancestral source gene by an average $K_S < 0.15$. This threshold corresponds roughly to duplications arising after divergence from the rodent lineage, either by recent gene duplication or perhaps recent gene conversion of older duplications (see Methods in Supplementary Information). A total of 1,183 genes exhibit such divergence from a neighbouring gene (see Methods in Supplementary Information). These genes often fall within larger clusters of paralogous genes including genes with greater divergence and reflecting older duplications. These clusters contain ~3,300 genes, and those having at least five genes involved in recent duplication events are shown in Table 4. Analysis of phylogenetic trees containing the related human and mouse genes confirms that the genes are more closely related within each species than between the two species in nearly all cases (97%), as would be expected for genes arising by duplication after the divergence of the human and rodent lineages.

The recent duplications are enriched in genes with immune and olfactory function, as well as those likely to be involved in reproductive functions. For example, the gene families encoding the pregnancy-specific beta-1-glycoprotein and choriongonadotropin beta proteins may be involved in the extended gestational period in the human lineage; the latter family is known to have expanded recently within the catarrhine primate lineage⁵⁷. Another example is the family of cancer/testis (CT) antigen genes, which are normally expressed in the testis and are highly expressed in carcinomas⁵⁸.

The distribution of K_S values (Fig. 7) for recent duplications shows a striking excess of genes with strong similarity ($K_S \leq 0.015$), corresponding to recent events occurring ~3–4 million years ago.

There are several possible explanations for this peak. First, it may reflect a true explosion in the rate of gene duplication in the primate lineage. (The primate lineage does show an increase in the rate of dispersed segmental duplication, although it is less extreme; the rate of local duplication will need to be carefully evaluated in comparative studies.) Second, it may partly reflect on the ongoing process of gene conversion of older gene duplication events. However, we offer a third explanation: the peak primarily reflects the transient of duplicated genes that are too young relative to the characteristic time of deletion. If so, most of these new genes are destined to be culled due to lack of functional benefit. In contrast to the first explanation, this would predict that a similar peak would be seen in most mammals.

Gene death in the human lineage

Gene death is another phenomenon that sheds light on lineage-specific evolution, but which was difficult to analyse with the earlier draft sequence. To study gene death, we scanned the genome sequence for recently arising non-processed pseudogenes—that is, nearly intact human genes that appear to have recently acquired an inactivating mutation. Specifically, we examined genomic intervals bounded at each end by two consecutive genes, with each belonging to a 1:1:1 orthology triplet in the human, mouse and rat genomes and the interval containing at most 50 genes (see Methods in Supplementary Information). We then examined the Ensembl gene predictions in the corresponding intervals of the three genomes and identified instances in which the mouse and rat genomes contained 1:1 orthologues, but the human genome appeared to contain no predicted orthologous gene. In each instance, the rodent genes were aligned to the corresponding human genomic interval to look for clear evidence of a human pseudogene—that is, a highly similar sequence containing one or more inactivating mutations in its genomic sequence (see Methods in Supplementary Information). We also required that the inactivating mutation was present in any human mRNA sequences corresponding to the locus. (This analysis excludes many older pseudogenes that do not show sufficient similarity to the rodent homologues because they have substantially degenerated.)

A total of 37 candidate pseudogenes were identified, with an average of 0.8 premature stop codons and 1.6 frameshifts (Supplementary Table 1). (Similar analyses performed on the draft sequence yielded a much larger list, including many apparent inactivating mutations that were errors and were corrected in the current sequence.) We carefully examined these candidates to confirm that they did not reflect errors in the current genome sequence (by resequencing or examination of an independently finished clone) and to determine their evolutionary origin (by resequencing in a panel of 24 diverse humans and comparison with a draft sequence of the chimpanzee genome). Complete experimental data could be obtained for 34 cases. The identification of a pseudogene was confirmed in 33 of the 34 cases; one case was due to an error in the current sequence (Table 5). The 19 pseudogenes with two or more inactivating mutations were all found to be pseudogenes in chimpanzee as well. The 14 pseudogenes with exactly one inactivating mutation fell into the following three classes: eight pseudogenes shared with chimpanzee; five pseudogenes fixed in the human population but functional genes in the chimpanzee; and one pseudogene that is a segregating polymorphism in the human population. (In 20 cases, the inactivating mutation occurs in the final or only exon. Although this could in principle be compatible with a functional gene, the truncation removes a functionally important domain in all but one case.)

Of the 32 pseudogenes fixed in the human population, 10 are derived from olfactory receptors. Olfactory receptors thus occur prominently in both birth and death analyses, indicating a dynamic expansion and contraction of this large gene family; the net effect has been an overall significant decrease in the number of functional

olfactory receptors in humans compared with rodents^{59,60}. The remaining 22 recent pseudogenes include a wide variety, such as genes homologous to a cationic amino-acid transporter, a serine-threonine kinase, a calreticulin, a putative G-protein coupled receptor and a cystatin.

Discussion

The Human Genome Project marked a new approach in biomedical research, one in which the scientific community came together to characterize systematically a large domain of important biological knowledge. Because the precise scientific plan and the feasible degree of accuracy and completeness were unclear at the outset, the sequencing of the human genome proceeded in phases: a preliminary phase that developed and refined key approaches; a draft phase that yielded ~90% of the information (albeit in imperfect form); and a finishing phase reported here that yielded ~99% in high-quality form. Notably, the finishing phase required roughly equal resources of time and expense as the draft phase.

The euchromatic portion of the human genome is still not complete, with ~1% still to be determined. The issue is no longer scale, but rather the need for new approaches to understand and resolve these recalcitrant segments. Continuing efforts should be devoted towards the eventual goal of complete closure. Nonetheless, the euchromatic human genome can now be regarded as effectively known. The accuracy and completeness of the current near-complete human genome sequence has important consequences for biomedical research. It allows systematic searches for the causes of disease—for example, to find all key heritable factors predisposing to diabetes or somatic mutations underlying breast cancer—with confidence that little can escape detection. It facilitates experimental tools to recognize cellular components—for example, detectors for mRNAs based on specific oligonucleotide probes or mass-spectrometric identification of proteins based on specific peptide sequences—with confidence that these features provide a unique signature. It allows sophisticated computational analyses—for example, to study genome structure and evolution—with confidence that subtle results will not be swamped or swayed by noisy data. At a practical level, it eliminates tedious confirmatory work by researchers, who can now rely on highly accurate information. At a conceptual level, the near-complete picture makes it reasonable for the first time to contemplate systems approaches to cellular circuitry, without fear that major components are missing.

The HGP provides an essential foundation for the sequencing and analysis of additional large genomes. With the experience gained from the human genome, it has already become scientifically and economically feasible to produce draft genome sequence from many vertebrates, which will be a crucial tool for identifying the functional elements in the human genome through comparative analysis. Ultimately, we believe that such projects should aim higher to produce genome sequence with even greater accuracy and completeness. This will require digesting the diverse experience from the finishing phase of human sequencing and selecting a subset of techniques that can be most efficiently streamlined and scaled up to improve accuracy and completeness of genome sequence. A good example is the systematic closure of gaps by primer-directed walking on fosmid templates covering each gap, which may be able to close the vast majority of gaps in a draft sequence in an automated fashion.

More generally, the HGP demonstrates the tremendous potential value of coordinated projects to create community resources to propel biomedical research. Key challenges that lie ahead⁶¹ include: (1) systematic identification of all genetic polymorphisms carried in the human population, to facilitate the study of their association with disease; this will require comprehensive study of hundreds to thousands of human genomes. (2) Systematic identification of all functional elements in the human genome, including genes, proteins, regulatory controls and structure elements; this will require

comparative analysis with many additional mammalian genomes and systematic application of diverse experimental techniques. (3) Systematic identification of all the 'modules' in which genes and proteins function together; this will require comprehensive study and improved interpretation of expression, localization and interaction in a temporal and spatial context. Absolute completeness will be elusive but, as with the HGP, obtaining the substantial majority of the information will greatly accelerate the pace of biomedical research in thousands of laboratories. □

Received 29 July; accepted 7 September 2004; doi:10.1038/nature03001.

1. NIH/CEPH Collaborative Mapping Group. A comprehensive genetic linkage map of the human genome. *Science* **258**, 67–86 (1992).
2. Gyapay, G. *et al.* The 1993–94 Genethon human genetic linkage map. *Nature Genet.* **7**, 246–339 (1994).
3. Murray, J. C. *et al.* A comprehensive human linkage map with centimorgan density. *Science* **265**, 2049–2054 (1994).
4. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
5. Hudson, T. J. *et al.* An STS-based map of the human genome. *Science* **270**, 1945–1954 (1995).
6. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
7. International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
8. Dietrich, W. F. *et al.* A comprehensive genetic map of the mouse genome. *Nature* **380**, 149–152 (1996).
9. Gregory, S. G. *et al.* A physical map of the mouse genome. *Nature* **418**, 743–750 (2002).
10. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
11. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1457 (1997).
12. Goffeau, A. *et al.* Life with 6,000 genes. *Science* **274**, 546–567 (1996).
13. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282**, 2012–2018 (1998).
14. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
15. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
16. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
17. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **401**, 489–495 (1999).
18. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
19. Celniker, S. E. *et al.* Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol.* **3**, 0079.1–0079.14 (2002).
20. Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
21. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
22. Heilig, R. *et al.* The DNA sequence and analysis of human chromosome 14. *Nature* **421**, 601–607 (2003).
23. Skaletsky, H. *et al.* The male-specific regions of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
24. Hillier, L. W. *et al.* The DNA sequence of human chromosome 7. *Nature* **424**, 157–164 (2003).
25. Mungall, A. J. *et al.* The DNA sequence and analysis of human chromosome 6. *Nature* **425**, 805–811 (2003).
26. Dunham, A. *et al.* The DNA sequence and analysis of human chromosome 13. *Nature* **428**, 522–528 (2004).
27. Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535 (2004).
28. Humphray, S. J. *et al.* DNA sequence and analysis of human chromosome 9. *Nature* **429**, 369–374 (2004).
29. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 10. *Nature* **429**, 375–381 (2004).
30. Schmutz, J. *et al.* The DNA sequence and comparative analysis of human chromosome 5. *Nature* **431**, 268–274 (2004).
31. Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* **9**, 1–4 (1999).
32. Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**, 365–368 (2004).
33. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence Project: update and current status. *Nucleic Acids Res.* **31**, 34–37 (2003).
34. Strausberg, R. L. *et al.* Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl Acad. Sci. USA* **99**, 16899–16903 (2002).
35. Furey, T. S. *et al.* Analysis of human mRNAs with the reference genome sequence reveals potential errors, polymorphisms, and RNA editing. *Genome Res.* (in the press).
36. Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T. & Olson, M. V. Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast artificial chromosome vector. *Proc. Natl Acad. Sci. USA* **86**, 6240–6244 (1989).
37. Eichler, E. E., Clark, R. A. & She, X. An assessment of the sequence gaps: unfinished business in a finished human genome. *Nature Rev. Genet.* **5**, 345–354 (2004).
38. Lai, Z. *et al.* A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nature Genet.* **23**, 309–313 (1999).
39. She, X. *et al.* The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**, 857–864 (2004).
40. Rudd, M. K. & Willard, H. F. Analysis of the centromeric regions of the human genome assembly. *Trends Genet.* (in the press).
41. Nilsson, M. *et al.* Padlock probes reveal single-nucleotide differences, parent of origin and *in situ*

- distribution of centromeric sequences in human chromosomes 13 and 21. *Nature Genet.* **16**, 252–255 (1997).
42. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
43. Johnson, M. E. *et al.* Positive selection of a novel gene family during the emergence of humans and great apes. *Nature* **413**, 514–519 (2001).
44. Bailey, J. A., Church, D. M., Ventura, M., Rocchi, M. & Eichler, E. E. Analysis of segmental duplications and genome assembly in the mouse. *Genome Res.* **14**, 789–801 (2004).
45. Tuzun, E., Bailey, J. A. & Eichler, E. E. Recent segmental duplications in the working draft assembly of the brown Norway Rat. *Genome Res.* **14**, 493–506 (2004).
46. Horvath, J. E., Bailey, J. A., Locke, D. P. & Eichler, E. E. Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum. Mol. Genet.* **10**, 2215–2223 (2001).
47. Collins, J. E. *et al.* Re-evaluating human gene annotation: a second-generation analysis of chromosome 22. *Genome Res.* **13**, 27–36 (2003).
48. Cliften, P. F. *et al.* Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res.* **11**, 1175–1186 (2001).
49. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
50. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
51. Roest Crolius, H. *et al.* Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature Genet.* **25**, 235–238 (2000).
52. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
53. Storz, G. An expanding universe of noncoding RNAs. *Science* **296**, 1260–1263 (2002).
54. Szymanski, M., Erdmann, V. A. & Barciszewski, J. Noncoding regulatory RNAs database. *Nucleic Acids Res.* **31**, 429–431 (2003).
55. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919 (2002).
56. Torrents, D., Suyama, M., Zdobnov, E. & Bork, P. A genome-wide survey of human pseudogenes. *Genome Res.* **13**, 2559–2567 (2003).
57. Maston, G. A. & Ruvolo, M. Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol. Biol. Evol.* **19**, 320–355 (2002).
58. Scanlan, M. J., Gure, A. O., Jungbluth, A. A., Old, L. J. & Chen, Y.-T. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol. Rev.* **188**, 22–32 (2002).
59. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. The complete human olfactory subgenome. *Genome Res.* **11**, 685–702 (2001).
60. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
61. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
62. Lee, C., Weverick, R., Fisher, B. B., Furguson-Smith, M. A. & Lin, C. C. Human centromeric DNAs. *Hum. Genet.* **100**, 291–304 (1997).
63. Morton, N. E. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476 (1991).
64. Madan, K. & Bobrow, M. Structural variation in chromosome no. 9. *Ann. Genet.* **17**, 81–86 (1974).
65. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
66. Bailey, J. A. *et al.* Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.* **70**, 83–100 (2002).
67. Loftus, B. J. *et al.* Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics* **60**, 295–308 (1999).
68. Gordon, D., Desmarais, C. & Green, P. Automated finishing with autofinish. *Genome Res.* **11**, 614–625 (2001).
69. Istrail, S. *et al.* Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl Acad. Sci. USA* **101**, 1916–1921 (2004).
70. McMurray, A. A., Sulston, J. E. & Quail, M. A. Short-insert libraries as a method of problem solving in genome sequencing. *Genome Res.* **8**, 562–566 (1998).
71. Heiner, C. R., Hunkapiller, K. L., Chen, S. M., Glass, J. I. & Chen, E. Y. Sequencing multimegabase-template DNA with BigDye terminator chemistry. *Genome Res.* **8**, 557–561 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Leja for graphic design and production of the figures. We would also like to thank the many dedicated support staff at the sequencing centres and funding agencies. In addition to finished sequence produced by the IHGSC between the time of publication of the draft human genome and April 2003, Build 35 contains some significant published finished sequence from other centres as listed in Table 1—for this we would like to acknowledge M. Adams, B. Roe and G. Evans. Build 35 also contains a small number of individual finished deposited accessions from a variety of other groups, which may or may not have been published; we would like to acknowledge all of this work. We acknowledge G. Sisk for help in preparing the manuscript. This work was supported by The Wellcome Trust; The US National Institutes of Health; The US Department of Energy; The Ministry of Education, Culture, Sports, Science and Technology, Japan; The Federal German Ministry of Education, Research, and Technology; Projektträger Biologie, Energie, Umwelt des BMBF und BMWt; the Max-Planck-Society; Deutsche Forschungsgemeinschaft; Thüringer Ministerium für Wissenschaft, Forschung, und Kunst; The Medical Research Council (UK); European Commission, Directorate Science, Research and Development; Chinese Academy of Sciences, Ministry of Science and Technology, National Natural Science Foundation of China.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F. S. Collins (fc23a@nih.gov), E. S. Lander (lander@broad.mit.edu), J. Rogers (jrh@sanger.ac.uk) or R. H. Waterston (waterston@gs.washington.edu). The sequence described here has been deposited in public databases, with the 24 human chromosomes having accession numbers NC000001 to NC000024.

Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype

Olivier Jaillon¹, Jean-Marc Aury¹, Frédéric Brunet², Jean-Louis Petit¹, Nicole Stange-Thomann³, Evan Mauceli³, Laurence Bouneau¹, Cécile Fischer¹, Catherine Ozouf-Costaz⁴, Alain Bernot¹, Sophie Nicaud¹, David Jaffe³, Sheila Fisher³, Georges Lutfalla⁵, Carole Dossat¹, Béatrice Segurens¹, Corinne Dasilva¹, Marcel Salanoubat¹, Michael Levy¹, Nathalie Boudet¹, Sergi Castellano⁶, Véronique Anthouard¹, Claire Jubin¹, Vanina Castelli¹, Michael Katinka¹, Benoît Vacherie¹, Christian Biémont⁷, Zineb Skalli¹, Laurence Cattolico¹, Julie Poulain¹, Véronique de Berardinis¹, Corinne Cruaud¹, Simone Duprat¹, Philippe Brottier¹, Jean-Pierre Coutanceau⁴, Jérôme Gouzy⁸, Genis Parra⁶, Guillaume Lardier¹, Charles Chapple⁶, Kevin J. McKernan⁹, Paul McEwan⁹, Stephanie Bosak⁹, Manolis Kellis³, Jean-Nicolas Volf¹⁰, Roderic Guigó⁶, Michael C. Zody³, Jill Mesirov³, Kerstin Lindblad-Toh³, Bruce Birren³, Chad Nusbaum³, Daniel Kahn⁸, Marc Robinson-Rechavi², Vincent Laudet², Vincent Schachter¹, Francis Quétier¹, William Saurin¹, Claude Scarpelli¹, Patrick Wincker¹, Eric S. Lander^{3,11}, Jean Weissenbach¹ & Hugues Roest Crollius^{1*}

¹UMR 8030 Genoscope, CNRS and Université d'Evry, 2 rue Gaston Crémieux, 91057 Evry Cedex, France

²Laboratoire de Biologie Moléculaire de la Cellule, CNRS UMR 5161, INRA UMR 1237, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France

³Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02141, USA

⁴Muséum National d'Histoire Naturelle, Département Systématique et Evolution, Service de Systématique Moléculaire, CNRS IFR 101, 43 rue Cuvier, 75231 Paris, France

⁵Défenses Antivirales et Antitumorales, CNRS UMR 5124, 1919 route de Mende, 34293 Montpellier Cedex 5, France

⁶Grup de Recerca en Informàtica Biomèdica, IMIM-UPF and Programa de Bioinformàtica i Genòmica (CRG), Barcelona, Catalonia, Spain

⁷CNRS UMR 5558 Biométrie et Biologie Evolutive, Université Lyon 1, 69622 Villeurbanne, France

⁸INRA-CNRS Laboratoire des Interactions Plantes Micro-organismes, 31326 Castanet Tolosan Cedex, France

⁹Agencourt Bioscience Corporation, Massachusetts 01915, USA

¹⁰Biofuture Research Group, Evolutionary Fish Genomics, Physiologische Chemie I, Biozentrum, University of Wuerzburg, Am Hubland, D-97074 Wuerzburg, Germany

¹¹Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

* Present address: CNRS UMR8541, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France

***Tetraodon nigroviridis* is a freshwater puffer fish with the smallest known vertebrate genome. Here, we report a draft genome sequence with long-range linkage and substantial anchoring to the 21 *Tetraodon* chromosomes. Genome analysis provides a greatly improved fish gene catalogue, including identifying key genes previously thought to be absent in fish. Comparison with other vertebrates and a urochordate indicates that fish proteins have diverged markedly faster than their mammalian homologues. Comparison with the human genome suggests ~900 previously unannotated human genes. Analysis of the *Tetraodon* and human genomes shows that whole-genome duplication occurred in the teleost fish lineage, subsequent to its divergence from mammals. The analysis also makes it possible to infer the basic structure of the ancestral bony vertebrate genome, which was composed of 12 chromosomes, and to reconstruct much of the evolutionary history of ancient and recent chromosome rearrangements leading to the modern human karyotype.**

Access to entire genome sequences is revolutionizing our understanding of how genetic information is stored and organized in DNA, and how it has evolved over time. The sequence of a genome provides exquisite detail of the gene catalogue within a species, and the recent analysis of near-complete genome sequences of three mammals (human¹, mouse² and rat³) shows the acceleration in the search for causal links between genotype and phenotype, which can then be related to physiological, ecological and evolutionary observations. The partial sequence of the compact puffer fish *Takifugu rubripes* genome was obtained recently and this survey provided a preliminary catalogue of fish genes⁴. However, the *Takifugu* assembly is highly fragmented and as a result important questions could not be addressed.

Here, we describe and analyse the genome sequence of the freshwater puffer fish *Tetraodon nigroviridis* with long-range linkage and extensive anchoring to chromosomes. *Tetraodon* resembles *Takifugu* in that it possesses one of the smallest known vertebrate genomes, but as a popular aquarium fish it is readily available and is easily maintained in tap water (see Supplementary Notes for

naming conventions, natural habitat and phylogeny). The two puffer fish diverged from a common ancestor between 18–30 million years (Myr) ago and from the common ancestor with mammals about 450 Myr ago⁵. This long evolutionary distance provides a good contrast to distinguish conserved features from neutrally evolving DNA by sequence comparison. *Tetraodon* sequences in fact had an important role in providing a reliable estimate of the number of genes in the human genome⁶.

There has been a vigorous and unresolved debate as to whether a whole-genome duplication (WGD) occurred in the ray-finned fish (actinopterygians) lineage after its separation from tetrapods^{7–9}. By exploiting the extensive anchoring of the *Tetraodon* sequence to chromosomes, we provide a definitive answer to this question. The distribution of duplicated genes in the genome reveals a striking pattern of chromosome pairing, and the correspondence of orthologues with the human genome show precisely the signatures expected from an ancient WGD followed by a massive loss of duplicated genes.

Moreover, we find that relatively few interchromosomal

Table 1 Assembly statistics

Parameter	Number	N50 length (kb)	Size with gaps included (Mb)	Size with gaps excluded (Mb)	Longest (kb)	Percentage of the genome with gaps included
All contigs	49,609	16	312.4	312.4	258	91.9
All scaffolds	25,773	984	342.4	312.4	7,612	100.7
All ultracontigs	128	7,622	276.4	247.0	12,035	81.3
Mapped contigs	16,083	26	197.7	197.7	258	58.1
Mapped scaffolds	1,588	608	218.4	197.7	7,612	64.2
Mapped ultracontigs	39	8,701	219.7	197.7	12,035	64.6

rearrangements occurred in the *Tetraodon* lineage over several hundred million years after the WGD. This allows us to propose a karyotype of the ancestral bony vertebrate (Osteichthyes) composed of 12 chromosomes, and to uncover many unknown evolutionary breakpoints that occurred in the human genome in the past 450 Myr.

The *Tetraodon* genome sequence

Sequencing and assembly

The *Tetraodon* genome was sequenced using the whole-genome shotgun (WGS) approach. Random paired-end sequences providing 8.3-fold redundant coverage were produced at Genoscope (GSC) and the Broad Institute of MIT and Harvard (see Supplementary Table S11). From this, the assembly program Arachne^{10,11} constructed 49,609 contigs for a total of 312 megabases (Mb; Table 1), which it then connected into 25,773 scaffolds (or supercontigs) covering 342 Mb (including gaps; see Supplementary Information). Half of the assembly is in 102 scaffolds larger than 731 kilobases (kb; the N50 length) and the largest scaffold measures 7.6 Mb, the typical length of a *Tetraodon* chromosome arm.

We produced additional data to physically link scaffolds and anchor them to chromosomes. These data include probe hybridizations to arrayed bacterial artificial chromosome (BAC) libraries,

restriction digest fingerprints of BAC clones, additional linking clone sequence, alignment to available *Takifugu* sequence and two-colour fluorescence *in situ* hybridization (FISH) (see Supplementary Information). The impact of these additional mapping data was twofold: first, we could join 2,563 scaffolds in 128 'ultracontigs' that cover 81.3% of the assembly, and second, we were able to anchor the 39 ultracontigs among the largest (covering 64.6% of the assembly, with an N50 size of 8.7 Mb) to *Tetraodon* chromosomes (Fig. 1; see also Supplementary Table S12 and Supplementary Notes).

The accuracy of the assembly was experimentally tested and the inter-contig links found to be correct in >99% of cases. On the basis of a re-sequencing experiment, we estimate that the assembly covers >90% of the euchromatin of the *Tetraodon* genome (Supplementary Information). Finally, the overall genome size was directly measured by flow cytometry experiments on several fish; an average value of 340 Mb was obtained, consistent with the sequence assembly and smaller than the previously reported estimate of 350–400 Mb.

The *Tetraodon* draft sequence has roughly 60-fold greater con-

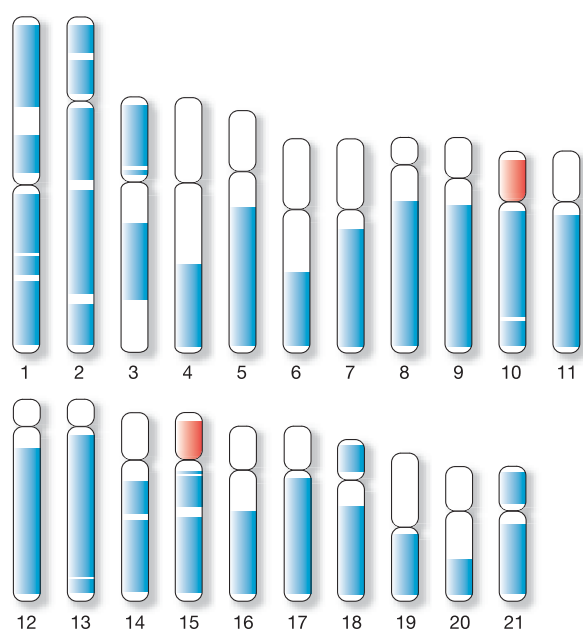


Figure 1 The *Tetraodon* genome is composed of 21 chromosomes. Red areas indicate the location of 5S and 28S ribosomal RNA gene arrays on chromosome 10 and chromosome 15, respectively. Many chromosomes are subtelocentric; that is, they only possess a very short heterochromatic arm. The extent of 39 sequence-based ultracontigs that cover about 64% of their length is shown in blue. In addition, approximately 16% of the genome is contained in another 89 ultracontigs that are not yet anchored on chromosomes, and the remaining 20% of the genome is in 23,210 smaller scaffolds.

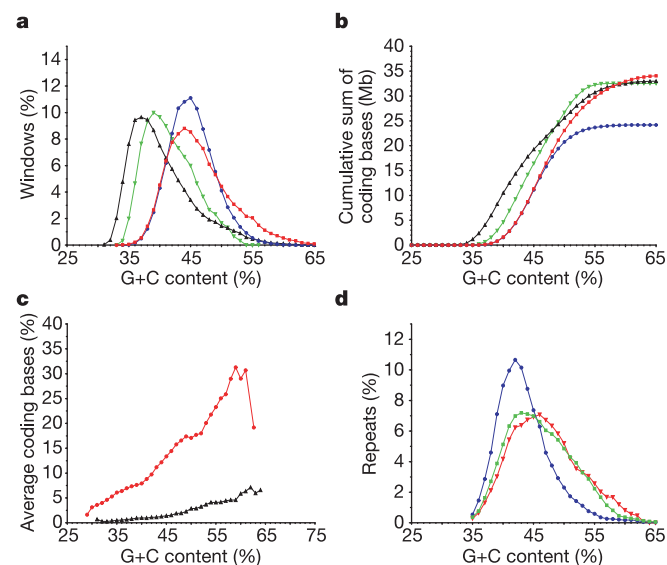


Figure 2 Distribution of the G + C content. **a**, Distribution in 5-kb non-overlapping windows across *Tetraodon* (red squares) and *Takifugu* (blue circles) scaffolds, and in 50-kb windows in human (black triangles) and mouse (green inverted triangles) chromosomes. Windows containing more than 25% ambiguous or unknown nucleotides (gaps) were excluded from the analysis. **b**, Cumulative sum of annotated coding bases in *Tetraodon* and *Takifugu* (5-kb non-overlapping windows) and human and mouse (50-kb windows) as a function of G + C content. **c**, In sharp contrast to *Takifugu*⁴ the density of genes increases with the G + C content (%) in *Tetraodon* (red circles) much more than in human (black triangles). **d**, The three major families of repeats in *Tetraodon* are not distributed uniformly in the genome: long terminal repeat (LTR) and LINE elements (red diamonds and green squares, respectively) concentrate in (G + C)-rich regions and SINE elements (blue circles) concentrate in (A + T)-rich regions. In contrast, the distribution of these elements is much more uniform in *Takifugu* (Supplementary Fig. S4).

Table 2 Comparison between *Tetraodon* and *Takifugu* annotations

Parameter	<i>Tetraodon</i>	<i>Takifugu</i> *	<i>Takifugu</i> †
Annotated genes	27,918	35,180	20,796
Annotated transcripts	27,918	38,510	33,003
Average number of coding exons per gene	6.9	4.3	8.6
Average number of UTR exons per gene	0.4	0†	0.07
Average gene size (bp)	4,778	2,754	6,547
Average CDS size (bp)	1,230	745	1,397
Average exon size (bp)	178	171	163
Number of annotated bases (Mb)			
Coding	33.9	26.1	29.1
UTR	2.4	0†	0.02

* *Takifugu* annotations are from Ensembl version 18.2.1.

† *Takifugu* annotations are from Ensembl version 23.2.1.

‡ *Takifugu* annotations from Ensembl version 18.2.1 do not include UTRs.

tinuity at the level of N50 ultracontig size than the *Takifugu* draft sequence (7.62 Mb versus 125 kb). Critically, the anchoring of the assembly provides a comprehensive view of a fish genome sequence organized in individual chromosomes.

Genome landscape

A consequence of the remarkably compact nature of the *Tetraodon* genome is that its G+C content is much higher than in the larger genomes of mammals. Although the G+C content is shifted markedly, it still shows the same asymmetric bell-shaped distribution with an excess of higher values as seen in human and mouse (Fig. 2a). (G+C)-rich regions tend to be gene-rich in mammals, and analysis of our data shows that this is also true for *Tetraodon* (Fig. 2b, c). The *Tetraodon* genome thus cannot be considered as a single homogeneous component but, as in mammals, it is a mosaic of relatively gene-rich and gene-poor regions.

Transposable elements are very rare in the *Tetraodon* genome^{12,13}: we estimate here that they do not exceed 4,000 copies; however, with 73 different types, they are richly represented (Supplementary Notes and Supplementary Table SI3). In sharp contrast, the human and mouse genomes contain only ~20 different types but are riddled with millions of transposable element copies. One of the intriguing features of the human genome is that the distribution of short interspersed nucleotide elements (SINEs) is biased towards (G+C)-rich regions, whereas long interspersed nucleotide elements (LINEs) favour (A+T)-rich regions. In *Tetraodon*, these preferences are precisely reverse: LINEs occur preferentially in (G+C)-rich

regions and SINEs in (A+T)-rich regions (Fig. 2d). The reason for these differences is not clear.

The *Tetraodon* genome shows certain striking differences from the previously reported *Takifugu* genome sequence. *Takifugu* contains eightfold more copies of transposable elements⁴ than *Tetraodon*, which may contribute to its slightly larger genome size (approximately 370 Mb; see Supplementary Information). More surprisingly, the G+C content of *Takifugu* does not show the characteristic asymmetry seen in mammals and in *Tetraodon* (Fig. 2a) nor the biases in SINE and LINE distribution (Supplementary Fig. S4). Why would the (G+C)-rich component be lacking in the *Takifugu* sequence, when this fraction is gene dense in mammals and in *Tetraodon*? This cannot be ascribed to transposable elements, which represent less than 5% of the assembly in both of these puffer fish species. One possible explanation is that the (G+C)-rich fraction exists in *Takifugu*, but was markedly under-represented as a result of aspects of the cloning, sequencing or assembly process. The fact that *Tetraodon* (G+C)-rich regions contain an excess of genes with no apparent orthologues in the *Takifugu* genome supports this hypothesis. Indeed, the *Tetraodon* genome appears to contain ~16.5% more coding exons than *Takifugu* (see below).

Tetraodon genes

Gene catalogue

The most prevalent features of the *Tetraodon* genome are protein-coding genes, which span 40% of the assembly. We constructed a catalogue of genes by adapting the GAZE¹⁴ computational framework (Supplementary Fig. S5) in order to combine three types of data: *Tetraodon* complementary DNA mapping, similarities to human, mouse and *Takifugu* proteins and genomes, and *ab initio* gene models (Supplementary Notes and Supplementary Tables SI4 and SI5).

The current *Tetraodon* catalogue is composed of 27,918 gene models, with 6.9 coding exons per gene on average (7.3 including untranslated regions (UTRs); Table 2). Assuming that fish and mammal genes possess similar gene structures, this suggests that some *Tetraodon* annotated genes are partial or fragmented because human and mouse genes respectively show 8.7 and 8.4 coding exons per gene². Adjusting the gene count for such fragmentation (by multiplying by 6.9/8.6) would yield an estimated gene count of 22,400 genes, whereas accounting for unsequenced regions of the genome might increase the estimate slightly further. Although such

Table 3 Comparative InterPro analysis of fish, mammal and urochordate proteomes

<i>Tetraodon</i>	<i>Takifugu</i>	Human	Mouse	<i>Ciona</i>	InterPro description
Actinopterygian-enriched					
61	78	22	21	48	Sodium:neurotransmitter symporter
33	29	11	13	33	Na ⁺ /solute symporter
21	16	8	7	6	Sodium/calcium exchanger membrane region
141	191	86	97	52	Collagen triple helix repeat
15	28	6	4	19	HAT dimerization
17	15	5	4	27	Peptidase M12A, astacin
3	4	0	0	1	Inosine/uridine-preferring nucleoside hydrolase
Sarcopterygian-enriched					
0	0	275	173	0	KRAB box
0	0	14	8	0	KRAB-related
3	0	25	29	0	High mobility group protein HMG14 and HMG17
0	0	9	95	0	Vomeroneasal receptor, type 1
0	0	13	21	0	Keratin, high sulphur B2 protein
0	0	3	3	0	Keratin, high-sulphur matrix protein
0	0	22	11	0	Mammalian taste receptor
0	0	11	9	0	Pancreatic RNase
0	0	7	8	0	β-Defensin
Vertebrate-enriched					
52	40	82	102	9	Histone core
252	253	240	228	88	Homeobox
62	56	80	55	9	Zn finger, B box
94	83	75	74	19	Zn-binding protein, LIM
65	56	70	135	17	HMG1/2 (high mobility group) box

Supplementary Table SI7 contains the top 100 InterPro domains in *Tetraodon*.

Table 4 Evolutionarily conserved regions between mammals and fish

Query genome	Target genome			
	<i>Tetraodon nigroviridis</i>	<i>Takifugu rubripes</i>	<i>Homo sapiens</i>	<i>Mus musculus</i>
<i>Tetraodon nigroviridis</i>	NA	ND	139,316	133,091
<i>Takifugu rubripes</i>	ND	NA	139,932	131,835
Combined fish	NA	NA	151,708	142,804
<i>Homo sapiens</i>	142,820	133,239	NA	ND
<i>Mus musculus</i>	140,407	129,996	ND	NA
Combined mammals	151,668	140,965	NA	NA

NA, not applicable; ND, not determined.

estimates are somewhat imprecise, it seems likely that *Tetraodon* has between 20,000–25,000 protein coding genes.

The *Tetraodon* gene catalogue appears to be the most complete so far for a fish, with coding exons and UTRs totalling ~36 Mb (~11% of the genome; Table 2). The *Takifugu* paper⁴ reported an estimate of 35,180 genes, but it did not account for a high degree of fragmentation (~4.3 exons per gene model). More recent, unpublished analyses have revised this number sharply downward (Table 2). The human and *Tetraodon* genomes have a similar distribution of exon sizes but markedly different distributions of intron size (Supplementary Fig. S6a). Although neither genome seems to tolerate introns below approximately 50–60 base pairs, *Tetraodon* has accumulated a much higher frequency of introns at this lower limit. Interestingly, this phenomenon is not uniform across the genome: there is an excess of genes with many small introns (Supplementary Fig. S6b), suggesting that intron sizes fluctuate in a regional fashion.

Proteome comparison between vertebrates

We examined in detail two gene families with unusual properties that represent challenges for automatic annotation procedures and have particular biological interest. The first is the family of selenoproteins, where the UGA codon encodes a rare cysteine analogue named selenocysteine (Sec) instead of signalling the end of translation as in all other genes¹⁵. We annotated 18 distinct families in *Tetraodon* based on similarities with the 19 protein families known in eukaryotes, and discovered a new selenoprotein that seems to be restricted to the actinopterygians among vertebrates and does not have a Cys counterpart in mammals. We also catalogued type I helical cytokines and their receptors (HCRI), a group of genes that were not found in the *Takifugu* genome⁴ because of their poor sequence conservation, leading to the hypothesis that fish may not possess this large family that includes hormones and interleukins. *Tetraodon*, in fact, contains 30 genes encoding HCRI with a typical D200 domain (Supplementary Fig. S7) and represents all families previously described in mammals¹⁶.

InterPro¹⁷ domains were annotated in protein sequences predicted in the *Tetraodon*, *Takifugu*, human, mouse and the urochordate *Ciona intestinalis*¹⁸ genome using InterProScan¹⁹. We did not identify major differences between fish and mammal InterPro families, except for a few striking cases (Table 3): (1) collagen molecules are much more diverse in fish than in mammals, with one *Tetraodon* gene containing 20 von Willebrand type A domains,

the largest number found so far in a single protein. (2) Some domains associated with sodium transport are noticeably enriched in fishes and *Ciona*, perhaps a reflection of their adaptation to saline aquatic environments that was lost in land vertebrates. (3) Purine nucleosidases usually involved in the recovery of purine nucleosides are more abundant in fish, including an allantoin pathway for purine degradation that is present in *Tetraodon* and absent in human. (4) Several hundred KRAB box transcriptional repressors involved in chromatin-mediated gene regulation exist in mammals and are totally absent in fish. (5) Proteins involved in general gene regulation are more abundant in vertebrates than in *Ciona*.

Protein annotation with gene ontology (GO) classifications²⁰ shows only subtle differences between fish and mammals, as was already observed between human and mouse². The largest differences between species are seen with the GO classification in molecular functions (Supplementary Fig. S9). Interestingly, the two puffer fish and *Ciona* often vary together, showing for instance a higher frequency of enzymatic and transporter functions, and a lower frequency of signal transducer and structural molecules than both mammals (human and mouse). These global observations are difficult to relate to evolutionary or physiological mechanisms but provide a framework to understand the emergence or decline of molecular functions in vertebrates.

Number of genes in mammals and teleosts

The total amount of coding sequence conserved between the two fish and the two mammalian genomes provides a measure of their respective coding capacity. The Exofish method⁶ is well suited to measure this, because it translates entire genomes in all six frames and identifies conserved coding regions (ecores) with a high specificity and independently of prior genome annotation (Table 4; see also Supplementary Information). The four vertebrate genomes contain remarkably similar numbers of ecores, apart from minor differences attributable to varying degrees of sequence completion. This suggests that they possess fairly similar numbers of genes. In fact, the gene count may be slightly less in mammals than in fish because the proportion of ecores corresponding to pseudogenes is higher in mammals²¹.

The human ecores can be used to search for previously unrecognized human genes. The discovery of new human genes is becoming an increasingly rare event, given the scale and intensity of international efforts to annotate the genome by systematic annotation pipelines and by human experts. Roughly 14,500 human ecores

Table 5 Rates of DNA evolution in vertebrates

Species	Total number of orthologues	Number of orthologues used	Average per cent identity (without gaps)	Observed number of substitutions per 4D site	Estimated amount of neutral evolution	Estimated rate of neutral evolution (sites per Myr)	K _a
Human–mouse	14,889	5,802	91.76	0.32	0.43	0.0057	0.05
<i>Tetraodon</i> – <i>Takifugu</i>	12,909	5,802	90.51	0.27	0.35	0.0146	0.06
<i>Tetraodon</i> –human	9,975	5,802	69.90	0.63	1.54*	–	0.24
<i>Tetraodon</i> –mouse	9,666	5,802	69.46	0.63	1.53*	–	0.25
<i>Takifugu</i> –human	9,143	5,802	70.05	0.63	1.52*	–	0.24
<i>Takifugu</i> –mouse	8,956	5,802	69.67	0.63	1.52*	–	0.25

* These values are saturated and cannot be considered reliable estimates.

conserved with *Tetraodon* sequences do not overlap any 'known' features (genes or pseudogenes) in the human genome. Using these as anchors for local gene identification using the GAZE program, we identified 904 novel human gene predictions. Of these, 63% are also supported by expressed sequence tag (EST) data (from human or other species) and 50% contain predicted InterPro protein domains (Supplementary Table SI9). The most convincing evidence supporting these gene predictions is that they are strongly enriched on chromosomes that have not yet been annotated by human experts (Supplementary Table SI10). The novel gene predictions have relatively small size (average coding sequence (CDS) of 469 bp), which may have caused them to be eliminated by systematic annotation procedures. They provide a rich resource to help complete the human gene catalogue.

Tetraodon gene evolution

We measured rates of sequence divergence between fish and mammals to estimate the relative speed with which functional and non-functional sequences evolve in these lineages. We used fourfold degenerate (4D) site substitutions in orthologous proteins as a proxy for neutral nucleotide mutations, an approach that has been shown to be robust across entire genomes². To optimize further the selection of sites used for comparison, we only considered the 5,802 proteins that are identified as orthologues in all pairwise comparisons between human, mouse, *Tetraodon* and *Takifugu*. The average neutral nucleotide substitution rate, inferred using the REV model^{22,23}, shows that the divergence between *Tetraodon* and *Takifugu* is about twice as fast per year as between human and mouse (Table 5), or between mouse and rat³.

We were interested to see whether this higher mutation rate is also seen in protein sequences. Pairwise comparison of all possible combinations of the 5,802 four-way orthologous proteins clearly indicates that proteins between the two puffer fish are more divergent than between the two mammals, despite the shorter evolutionary time that has elapsed (Fig. 3). This is confirmed by

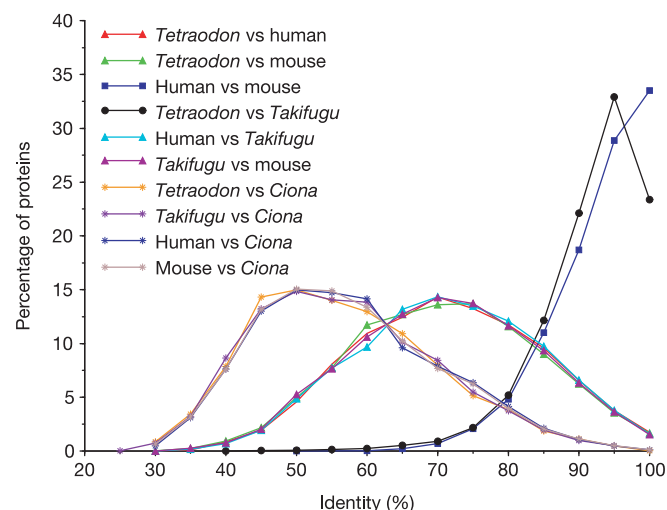


Figure 3 Distribution of the per cent identity between pairs of orthologous protein sets. Comparisons were performed with 2,289 proteins that are orthologous between the chordate *C. intestinalis* and all four vertebrates—*Tetraodon*, *Takifugu*, human and mouse (asterisks)—and with 5,802 proteins orthologous between all four vertebrates only, between fish and mammals (triangles) or between the two fish (circles), and between the two mammals (squares). As expected, all vertebrates show the same distribution profile compared to *Ciona* and both fish show the same distribution profile compared to mammals. Surprisingly, the distribution profile of the comparison between the two fish and between the two mammals is also very similar, despite the much shorter evolutionary time since the tetraodontiform radiation.

the fact that the average frequency of non-synonymous mutations (leading to an amino acid change, K_a) between *C. intestinalis* and human proteins is lower than between *Ciona* and *Tetraodon* (see Methods).

Independent of the overall rate of change, the ratio of non-synonymous to synonymous changes (K_a/K_s ratio) is much higher between the two puffer fish than between human and mouse (Supplementary Table SI11 and Supplementary Information), suggesting that protein evolution is proceeding more rapidly along the puffer fish lineage. The reasons for this faster tempo of protein change are unknown, although it is likely to be positively correlated with the higher rate of neutral mutation.

Genome evolution

Genome-wide sequence provides a rare opportunity to address key evolutionary questions in a global fashion, circumventing biases due to small sequence and gene samples. In this respect, the combination of long-range linkage in the *Tetraodon* sequence and its evolutionary divergence from the mammalian lineage at 450 Myr ago makes it possible to explore overall genome evolution in the vertebrate clade.

Evidence for whole-genome duplication

The occurrence of WGD in the ray-finned fish lineage is a hotly debated question due both to the cataclysmic nature of such an event and to the difficulty in establishing that it actually occurred^{24–26}.

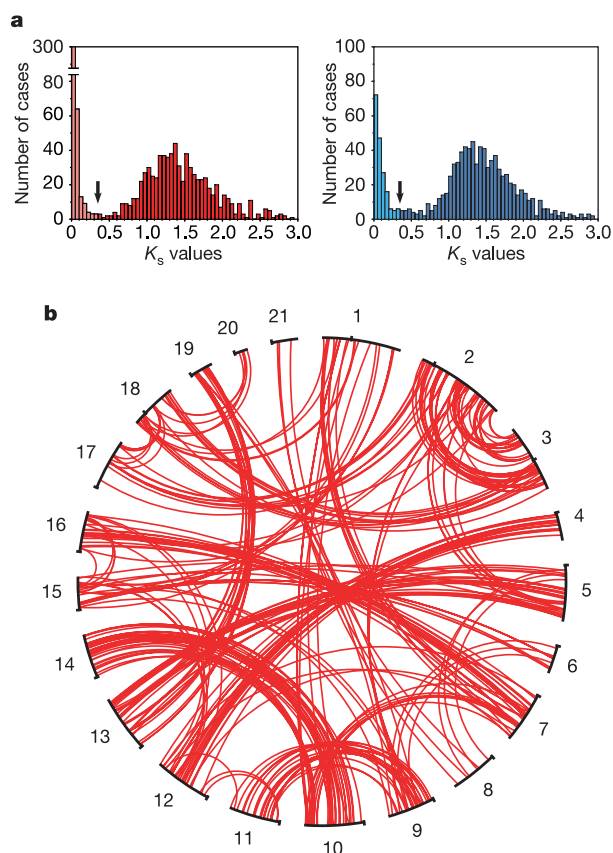


Figure 4 Genome duplication. **a**, Distribution of K_s values of duplicated genes in *Tetraodon* (left) and *Takifugu* (right) genomes. Duplicated genes broadly belong to two categories, depending on their K_s value being below or higher than 0.35 substitutions per site since the divergence between the two puffer fish (arrows). **b**, Global distribution of ancient duplicated genes ($K_s > 0.35$) in the *Tetraodon* genome. The 21 *Tetraodon* chromosomes are represented in a circle in numerical order and each line joins duplicated genes at their respective position on a given pair of chromosomes.

Definitive proof of WGD requires identifying certain distinctive signatures in long-range genome organization, which has previously been impossible to address with the data available.

It is expected that after WGD the resulting polyploid genome gradually returns to a diploid state through extensive gene deletion, with only a small proportion of duplicated copies ultimately

retained as sources of functional innovation²⁶. Paralogous chromosomes will thus each retain only a small subset of their initially common gene complement and then will be broken into smaller segments by genomic rearrangements. WGD will thus leave two distinctive signs for considerable periods before eventually fading.

The first distinctive sign is duplicated genes on paralogous chromosomes. In the absence of chromosomal rearrangement it would be simple to recognize two paralogous chromosomes arising from a WGD from the genome-wide distribution of duplicate genes: the chromosomes would each contain one member from many duplicated gene pairs occurring in the same order along their length. The difficulty is that this neat picture will eventually be blurred by interchromosomal rearrangement, which will disrupt the 1:1 correspondence between chromosomes, and intrachromosomal rearrangement, which will disrupt gene ordering along chromosomes.

We analysed the genome-wide distribution of duplicated gene pairs to see whether a strong correspondence between chromosomes could be detected. We identified 1,078 and 995 pairs of duplicated genes in the *Tetraodon* and *Takifugu* genomes, respectively, using conservative criteria (see Supplementary Information). On the basis of the frequencies of silent mutations (K_s) between copies, ~75% are ‘ancient’ duplications that arose before the *Tetraodon*–*Takifugu* speciation (Fig. 4a).

The chromosomal distribution of these ancient duplicates follows a striking pattern characteristic of a WGD. Genes on one chromosome segment have a strong tendency to possess duplicate copies on a single other chromosome (Fig. 4b). The correspondence is not a perfect 1:1 match owing to interchromosomal exchange, but it is vastly stronger than expected by chance (Supplementary Table S112). As expected from a WGD, all chromosomes are involved. Remarkably, some duplicate chromosome pairs such as *Tetraodon* chromosome 9 (Tni9) and Tni11 have remained largely undisturbed by chromosome translocations since the duplication event. In other cases, one chromosome has links to two or three others, suggestive of either fusion or fragmentation (for example, Tni13 matches Tni5 and Tni19).

The second distinctive sign, which is an even more powerful signature of genome duplication, comes from comparison with a related species carrying a genome that did not undergo the WGD. Such a comparison was recently used to prove the existence of an ancient WGD in the yeast *Saccharomyces cerevisiae* based on comparison with a second yeast species *Kluyveromyces waltii* that diverged before the WGD^{27,28}. Although two ancient paralogous regions typically retained only a few genes in common, they could be readily recognized because they showed a characteristic 2:1 mapping with interleaving; that is, they both showed conserved synteny and local order to the same region of the *K. waltii* genome with the *S. cerevisiae* genes interleaving in alternating stretches. Such regions were called blocks of DCS (doubly conserved synteny). Whereas the first distinctive sign of WGD depends only on a

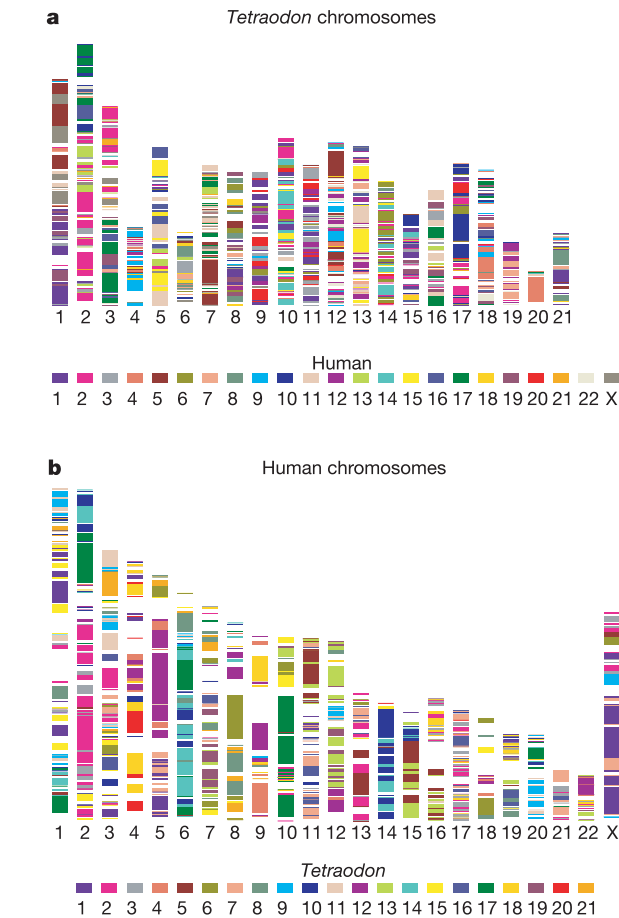


Figure 5 Synteny maps. **a**, For each *Tetraodon* chromosome, coloured segments represent conserved synteny with a particular human chromosome. Synteny is defined as groups of two or more *Tetraodon* genes that possess an orthologue on the same human chromosome, irrespective of orientation or order. *Tetraodon* chromosomes are not in descending order by size because of unequal sequence coverage. The entire map includes 5,518 orthologues in 900 syntenic segments. **b**, On the human genome the map is composed of 905 syntenic segments. See Supplementary Information for the synteny map between *Tetraodon* and mouse (Supplementary Fig. S11).

Table 6 Distribution of human orthologues on *Tetraodon* chromosomes listed by their ancestral chromosome of origin

	Ancestral chromosome											
	A	B	C	D	E	F	G	H	I	J	K	L
<i>Tetraodon</i> chromosome (copy 1)	4	17	2	2	5	13	7	1	1	10	9	6
Number of orthologues on copy 1	141	30	130	318	187	145	136	143	151	262	214	111
Percentage of orthologues on copy 1*	32.0	19.2	31.4	62.1	52.1	58.5	58.1	58.8	61.6	52.5	45.2	36.4
<i>Tetraodon</i> chromosome (copy 2)	12	18	3	3	13	19	16	7	15	14	11	8
Number of orthologues on copy 2	299	94	166	97	172	103	98	100	94	237	259	129
Percentage of orthologues on copy 2*	68.0	60.26	40.1	18.9	47.9	41.5	41.9	41.2	38.4	47.5	54.8	42.3
<i>Tetraodon</i> chromosome (copy 3)	–	20	18	17	–	–	–	–	–	–	–	21
Number of orthologues on copy 3	–	32	118	97	–	–	–	–	–	–	–	65
Percentage of orthologues on copy 3*	–	20.5	28.50	18.9	–	–	–	–	–	–	–	21.31

*Only orthologues that belong to syntenic groups are indicated here. For instance, ancestral chromosome A could be reconstructed with 141 *Tetraodon*–human orthologues belonging to *Tetraodon* chromosome 4 and 299 to chromosome 12.

minority of duplicated genes, the DCS signature considers all genes for which orthologues can be found in the related species.

We used 6,684 *Tetraodon* genes localized on individual chromosomes that possess an orthologue in either human or mouse to create a high-resolution synteny map (Fig. 5 and Supplementary Fig. S11, respectively). The map contains 900 syntenic groups composed of at least two consecutive genes (average 6.1; maximum 55) having orthologues on the same human chromosome; the syntenic groups include 76% of *Tetraodon*–human orthologues. The synteny map with mouse contains 1,011 syntenic groups, probably reflecting the higher degree of chromosomal rearrangement in the rodent lineage².

The synteny map typically associates two regions in *Tetraodon* with one region in human. Using precise criteria (see Methods) we defined DCS blocks for *Tetraodon* relative to human; in contrast to

the yeast study, strict conservation of gene order within DCSs was not required. Notably, most (79.6%) orthologous genes in syntenic groups can be assigned to 90 DCS blocks (Fig. 6). As in *S. cerevisiae*²⁷, we see the distinctive interleaving pattern expected from WGD followed by massive gene loss. Analysis of the interleaving pattern shows that the gene loss occurred through many small deletions in a balanced fashion over the two *Tetraodon* sister chromosomes (average balance 42% and 58% of retention; Supplementary Information); this is consistent with the results in yeast.

These two analyses provide definitive evidence that the *Tetraodon* genome underwent a WGD sometime after its divergence from the mammalian lineage. The first test used only the ~3% of genes that represent duplicated gene pairs retained from the WGD. The second test used the pattern of 2:1 mapping with interleaving involving ~80% of orthologues between *Tetraodon* and human.

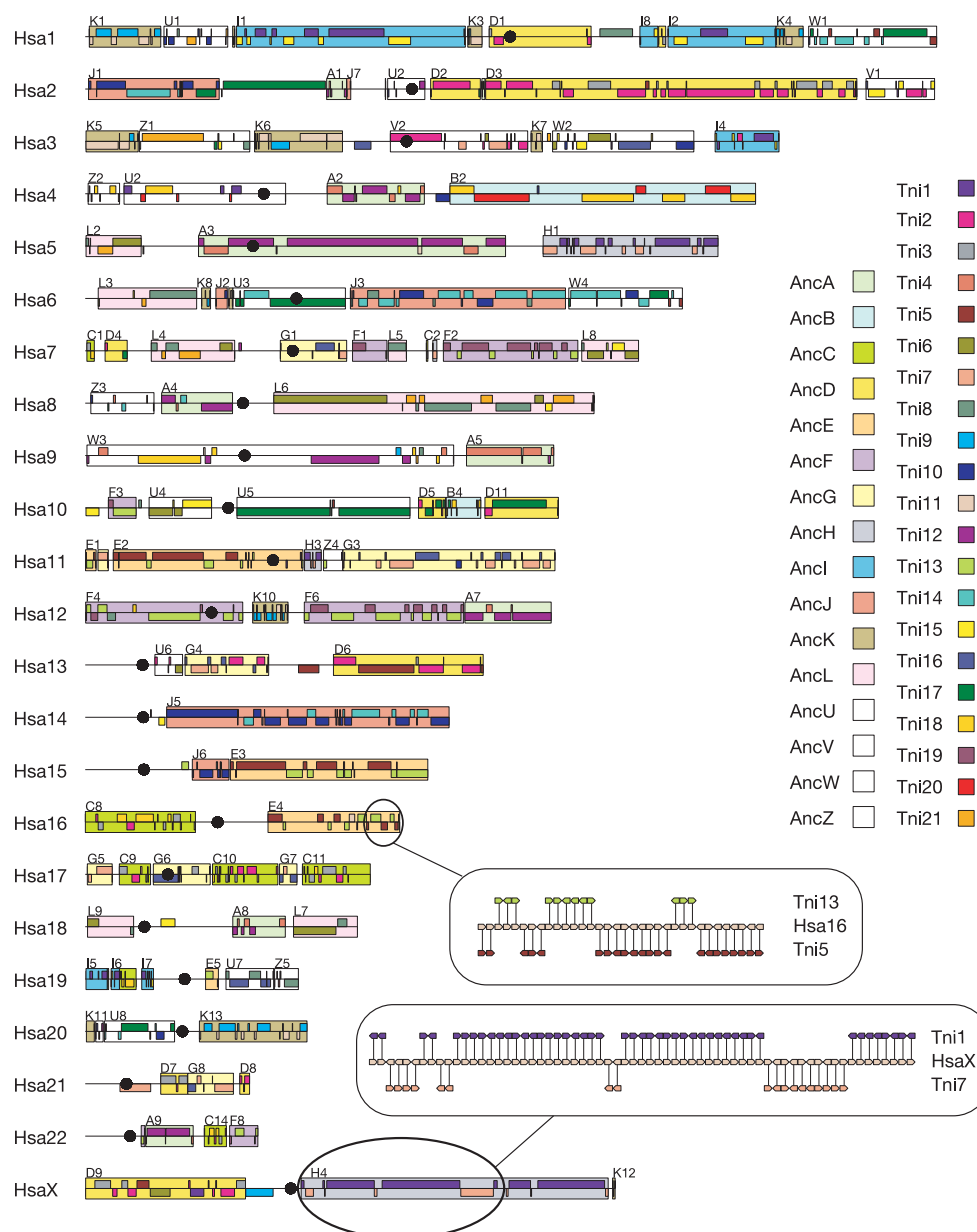


Figure 6 Duplicate mapping of human chromosomes reveals a whole-genome duplication in *Tetraodon*. Blocks of synteny along human chromosomes map to two (or three) *Tetraodon* chromosomes in an interleaving pattern. Small boxes represent groups of syntenic orthologous genes enclosed in larger boxes that define the boundaries of 110

DCS blocks. Black circles indicate human centromeres. A region of human chromosomes Xq and 16q are shown in detail with individual *Tetraodon* orthologous genes depicted on either side.

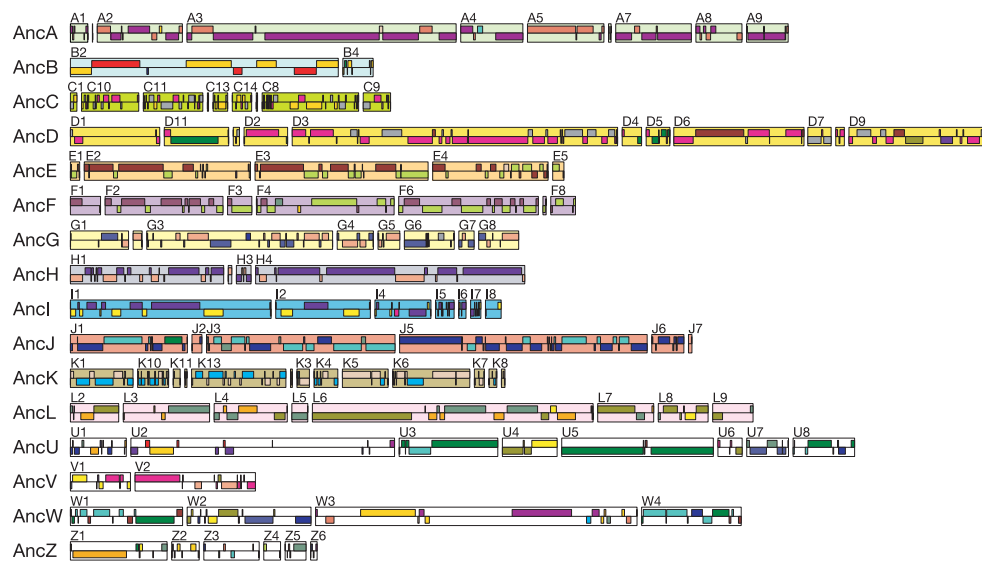


Figure 7 Composition of the ancestral osteichthyan genome. The 110 DCS blocks identified on the human genome are grouped according to their composition in terms of *Tetraodon* chromosomes, thus delineating 12 ancestral chromosomes containing 90 DCS blocks. The order of DCSs within an ancestral chromosome is arbitrary. The 20 blocks

denoted by the letters U, V, W and Z (Supplementary Information) could not be assigned to an ancestral chromosome because each has a unique composition, probably due to rearrangements in the human or *Tetraodon* genome. Colour codes are as in Fig. 6.

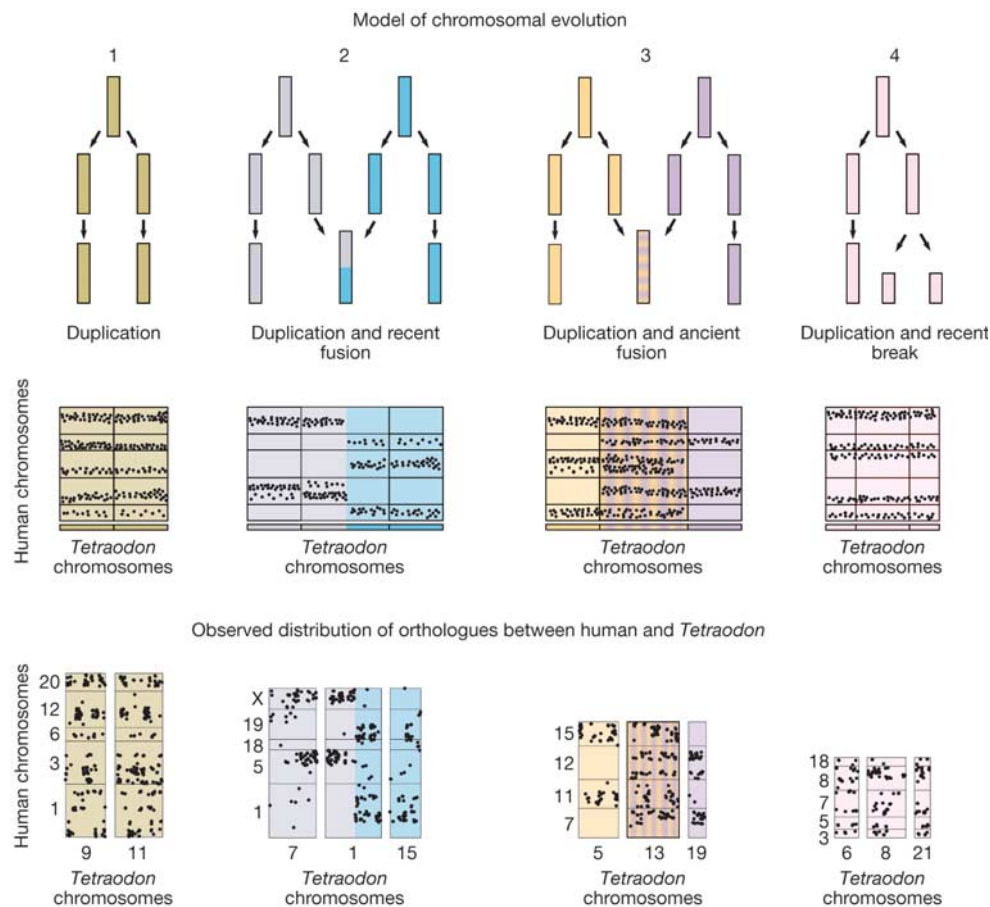


Figure 8 Reconstructing ancient genome rearrangements. Model of chromosome duplication followed by the four simplest chromosome rearrangements: (1) no rearrangement; (2) two different duplicate copies fused recently; (3) two different duplicate copies fused early after the duplication; (4) a duplicate chromosome fragmented very recently. In each model, the distribution of human orthologues from a given chromosomal region on two or three duplicate *Tetraodon* chromosomal regions is expected to be different (each dot is an orthologue, positioned in the human genome on the vertical axis and in the *Tetraodon* genome on the horizontal axis). The distinction

between early or late events follows the assumption that intrachromosomal shuffling progressively redistributes genes over a given chromosome. A recent fusion would thus bring together two sets of genes that appear compartmented on their respective segments, whereas an ancient fusion shows the same pattern except that genes have been redistributed over the length of the fused chromosome. It should be noted that a fifth case exists, consisting of a chromosome break early after duplication but it is not represented here. The lower panel shows excerpts of data illustrating the four types of event. The complete Oxford grid is shown in Supplementary Fig. SI12.

The presence of supernumerary HOX clusters in zebrafish⁷, *Tetraodon* (Fig. S8) and many other percomorphs²⁹ but not in the bichir *Polypterus senegalus*³⁰ indicates that the event has affected most teleosts but not all actinopterygians. This timing early in the teleost lineage is in agreement with recent evolutionary analyses in *Takifugu* that estimated the divergence time for most duplicated gene pairs at ~320–350 Myr ago^{31,32}.

The analyses above also shed light on the rate of intra- and interchromosomal exchange. The synteny analysis shows extensive syntenic segments in which gene content has been well preserved

but gene order has been extensively scrambled (striking examples include conserved synteny of *Tni20* with human chromosome 4q (Hsa4q) and *Tni1* with HsaXq); this is consistent with observations in zebrafish³³. The duplication analysis within *Tetraodon* also shows that the chromosomal correspondence of duplicated gene pairs has been extensively preserved, whereas local gene order has been largely scrambled. Both analyses thus indicate that a relatively high degree of intrachromosomal rearrangement and a relatively low degree of interchromosomal exchange have taken place in the *Tetraodon* lineage.

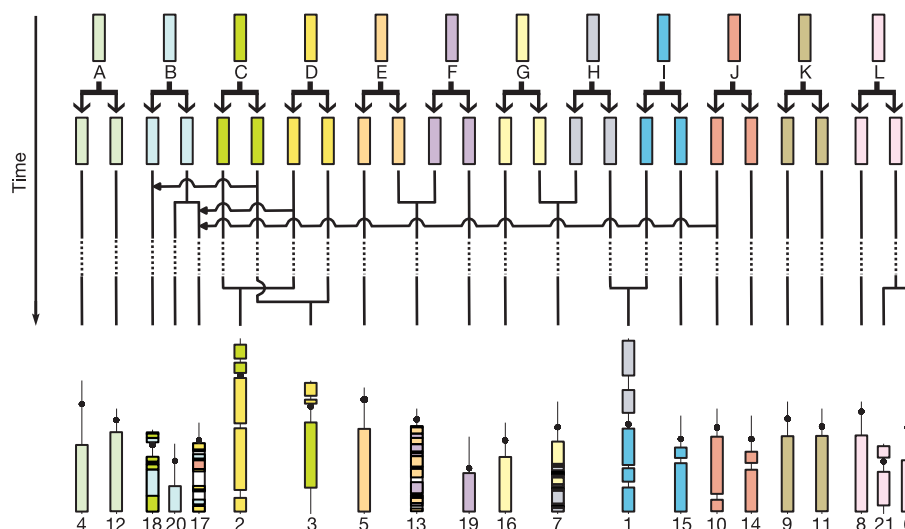


Figure 9 Model for the reconstruction of an ancestral bony vertebrate karyotype comprising 12 chromosomes, based on the pairing information provided by duplicated *Tetraodon* chromosomes showing interleaved patterns on human chromosomes. The ten major rearrangements (two ancient fusions, three recent fusions, one ancient and one recent fission, and three ancient translocations) are deduced by fitting the distribution of

orthologues to the four simple theoretical models of chromosome evolution. The order between events is arbitrary although the approximate timeline differentiates between ancient and recent events respectively before and after the dashed line. Arrowheads point to the direction of three ancient translocations.

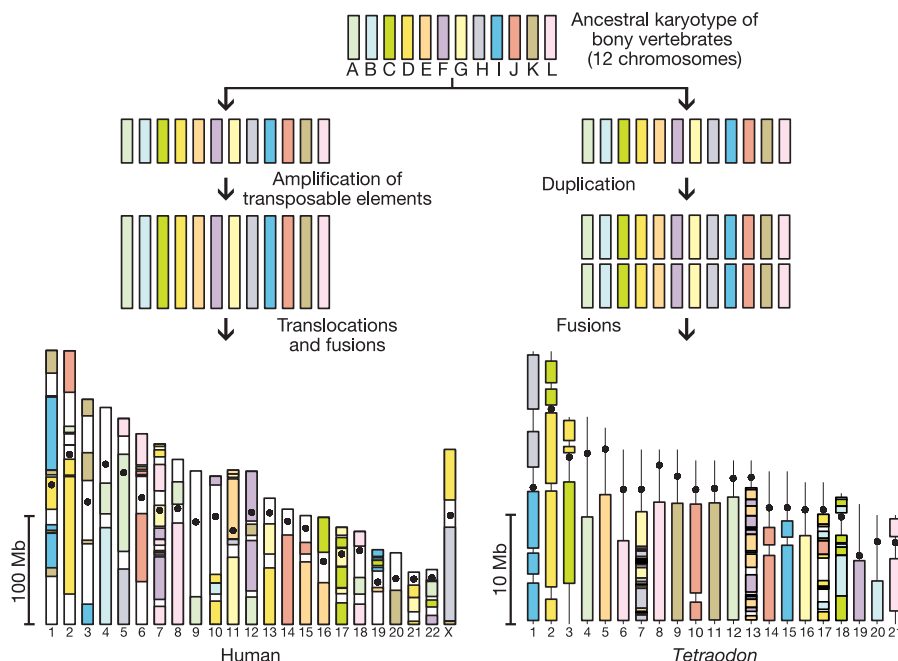


Figure 10 Proposed model for the distribution of ancestral chromosome segments in the human and the *Tetraodon* genomes. The composition of *Tetraodon* chromosomes is based on their duplication pattern (Fig. 9), whereas the composition of human chromosomes is based on the distribution of orthologues of *Tetraodon* genes (Fig. 6). A

vertical line in *Tetraodon* chromosomes denotes regions where sequence has not yet been assigned. With 90 blocks in human compared with 44 in *Tetraodon*, the complexity of the mosaic of ancestral segments in human chromosomes underlines the higher frequency of rearrangements to which they were submitted during the same evolutionary period.

Ancestral genome of bony vertebrates

We then sought to use the correspondence between the *Tetraodon* and human genomes to attempt to reconstruct the karyotype of their osteichthyan (bony vertebrate) ancestor. The DCS blocks define *Tetraodon* regions that arose from duplication of a common ancestral region. Notably, the DCS blocks largely fall into 12 simple patterns: eight cases involving the interleaving of two current *Tetraodon* chromosomes and four cases involving three current *Tetraodon* chromosomes (Fig. 7 and Table 6). The first group represents cases in which the ancestral chromosomes have remained largely untouched by interchromosomal exchange; the second group represents cases in which one major translocation has occurred.

The distribution of *Tetraodon* orthologues in the human genome (shown as an Oxford grid in Supplementary Fig. S12) provides a detailed record that can be used to partially reconstruct the history of rearrangements in both lineages. We considered the expected distribution resulting from various types of interchromosomal rearrangements, assuming a relatively high degree of intrachromosomal shuffling (Fig. 8; see also Supplementary Information). We found that only ten large-scale interchromosomal events suffice to largely explain the data, connecting an ancestral vertebrate karyotype of 12 chromosomes to the modern *Tetraodon* genome of 21 chromosomes (Fig. 9). Eleven of the *Tetraodon* chromosomes appear to have undergone no major interchromosomal rearrangement. For example, 13 DCS blocks in human are composed of interleaved syntenic groups mapping to Tni9 and Tni11, which are presumed to be derived from a common ancestral chromosome denoted chromosome K (AncK; Fig. 7). The orthologue distribution between the two chromosomes (Fig. 8) confirms that they derive by duplication from AncK (Fig. 9). In a more complex case, Tni13 is systematically interleaved with Tni5 (AncE) or Tni19 (AncF), but Tni5 and Tni19 are never interleaved together; the orthologue distribution among the three chromosomes (Fig. 8) implies that the duplication partners of Tni5 and Tni19 fused soon after the WGD to give rise to Tni13 (Fig. 9). The overall model is consistent with a complete WGD, in that it accounts for all *Tetraodon* chromosomes.

Several lines of evidence support the historical reconstitution presented here. First, the pairing of *Tetraodon* chromosomes agrees with the independently derived distribution of duplicated genes in the genome (Fig. 4b). Second, centric fusions of the three largest chromosomes are consistent with cytogenetic studies³⁴, and the recent timing of the fusion leading to Tni1 is supported by cytogenetic studies showing its absence in *Takifugu*³⁵. Third, the modal value for the haploid number of chromosomes in teleosts is 24 (refs 36–38), consistent with a WGD of an ancestral genome composed of 12 chromosomes.

The analysis also sheds light on genome evolution in the human lineage, with the interleaving patterns on human chromosomes delineating the mosaic of ancestral segments in the human genome (Figs 6 and 10). The results are consistent with and extend several known cases of rearrangements in the human lineage. The model correctly shows the recent fusion of two primate chromosomes leading to Hsa2 (ref. 39) occurring at the junction between two ancestral segments (D2 and D3; Fig. 6) in 2q13.2–2q14.1. It shows HsaXp and HsaXq to be of different origins (corresponding to AncD and AncH, respectively), consistent with the fact that HsaXp is known to be absent in non-placental mammals⁴⁰. The map indicates that most of HsaXq and Hsa5q were once part of the same chromosome, but that the tip of HsaXq (Xq28) originates from a different ancestral segment and is thus a later addition. Some pairs of human chromosomes show similar or identical compositions, suggesting that they derived by fission from the same ancestral chromosome, with examples being Hsa13–Hsa21 and Hsa12–Hsa22; the latter case is consistent with cytogenetic studies showing that a fission occurred in the primate lineage⁴¹.

The results show a major difference in the evolutionary forces shaping the *Tetraodon* and the human genomes (Fig. 10). Whereas 11 *Tetraodon* chromosomes did not undergo interchromosomal exchange over 450 Myr, only one human chromosome (Hsa14) was similarly undisturbed. Hsa7 is an extreme case, with contributions from six ancestral chromosomes. A possible explanation for the difference may be the massive integration of transposable elements in the human genome. The presence of transposable elements may increase the overall frequency of chromosome breaks, as well as the likelihood that a chromosome break fails to disrupt a gene (by increasing the size of intergenic intervals). It will be interesting to see whether teleosts that carry many more transposable elements (such as zebrafish) show a higher frequency of interchromosomal exchanges.

Conclusion

The purpose of sequencing the *Tetraodon* genome was to use comparative analysis to illuminate the human genome in particular and vertebrate genomes in general. The *Tetraodon* sequence, which has been made freely available during the course of this project, has already had a major impact on human gene annotation. It has provided the first clear evidence of a sharply lower human gene count⁶ and has been used in the annotation of several human chromosomes^{42–45}. Here, we show that it suggests an additional ~900 predicted genes in the human genome. Given its compact size, the *Tetraodon* genome will probably also prove valuable in identifying key conserved regulatory features in intergenic and intronic regions.

In addition, the *Tetraodon* genome provides fundamental insight into genome evolution in the vertebrate lineage. First, the analysis here shows that *Tetraodon* is the descendant of an ancient WGD that most probably affected all teleosts. Together with the recent demonstration of an ancient WGD in the yeast lineage, this suggests that WGD followed by massive gene loss may be an extremely important mechanism for eukaryote genome evolution—perhaps because it allows for the neofunctionalization of entire pathways rather than simply individual genes. There remains a fierce debate about whether one or more earlier WGD events occurred in early vertebrate evolution^{25,46–50}, with no direct and conclusive evidence found so far^{51,52}. The examples of yeast and *Tetraodon* show that ultimate proof will probably best come from the sequence of a related non-duplicated species. An obvious candidate is amphioxus, as its non-duplicated status is supported by the presence of many single-copy genes (including one HOX cluster⁵³) instead of two or more in vertebrates, and it is among our closest non-vertebrate relatives based on anatomical and evolutionary observations.

Second, the remarkable preservation of the *Tetraodon* genome after WGD makes it possible to infer the history of vertebrate chromosome evolution. The model suggests that the ancestral vertebrate genome was comprised of 12 chromosomes, was compact, and contained not significantly fewer genes than modern vertebrates (inasmuch as the WGD and subsequent massive gene loss resulted in only a tiny fraction of duplicate genes being retained). The explosion of transposable elements in the mammalian lineage, subsequent to divergence from the teleost lineage, may have provided the conditions for increased interchromosomal rearrangements in mammals; in contrast, the *Tetraodon* genome underwent much less interchromosomal rearrangement.

With the availability of additional vertebrate genomes (dog, marsupial, chicken, medaka, zebrafish and frog are underway), it will be possible to explore intermediate nodes such as the last common ancestor of amniotes, of sarcopterygians and of actinopterygians, and to gain an increasingly clearer picture of the early vertebrate ancestor. Because the early vertebrate genome is ‘closer’ to current invertebrates, this should in turn facilitate comparison between vertebrate and invertebrate evolution. □

Methods

Sequencing, assembly and data access

Sequencing was performed as described previously for Genoscope⁵⁴ and the Broad Institute⁵². Approximately 4.2 million plasmid reads were cloned and sequenced from DNA extracted from two wild *Tetraodon* fish and passed extensive checks for quality and source, representing approximately 8.3-fold sequence coverage of the *Tetraodon* genome. To alleviate problems due to polymorphism, the assembly proceeded in four stages: (1) reads from a single fish were assembled by Arachne as described previously^{10,11}; (2) reads from the second individual were added to increase sequencing depth; (3) scaffolds were constructed using plasmid and BAC paired reads; and (4) contigs from a separate assembly combining both individuals were added if they did not overlap with the first assembly. The final assembly can be downloaded from the EMBL/GenBank/DBJ databases under accession number CAAE01000000. Full-length *Tetraodon* cDNAs have been submitted under accession numbers CR631133–CR735083. Ultracontigs organized in chromosomes are available from <http://www.genoscope.org/tetraodon>. This site also contains an annotation browser and further information on the project.

Gene annotation

Protein-coding genes were predicted by combining three types of information: alignments with proteins and genomic DNA from other species, *Tetraodon* cDNAs, and *ab initio* models. All alignments with genomic DNA from human and mouse were performed with Exofish as described previously⁶, whereas a new Exofish method was developed to align *Takifugu* genomic DNA. Proteins predicted from human and mouse were also matched using Exofish and a selected subset was then aligned using Genewise. The integration of these data sources was performed with GAZE¹⁴. A specific GAZE automaton was designed, and parameters were adjusted on a training set of 184 manually annotated *Tetraodon* genes. See Supplementary Information for details.

Evolution of coding and non-coding DNA

To identify orthologous genes between human, mouse, *Tetraodon*, *Takifugu* and *Ciona*, their predicted proteomes were compared using the Smith–Waterman algorithm and reciprocal best matches were considered as orthologous genes between two species. However, only those genes that were reciprocal best matches between four or five species, and only sites that were aligned between the four or five genes, were further considered to compute the percentage identity, K_a , K_s and fourfold degenerate sites by the PBL method applying Kimura's two-parameter model^{55–57}. See Supplementary Information for details.

Genome duplication

A core set of *Tetraodon* duplicated genes was identified by an all-against-all comparison of *Tetraodon* predicted protein using Exofish. Only proteins that matched a single other protein by reciprocal best match were considered further and realigned by the Smith–Waterman algorithm to compute K_a and K_s values. Duplicates with a $K_s > 0.35$ (the amount of neutral substitution since the *Tetraodon*–*Takifugu* divergence) were considered 'ancient' and used to calculate P -values for chromosome pairing (Supplementary Table S12). Rules for classifying alternating patterns of syntenic groups along human chromosomes in DCS blocks included the following criteria: number of genes in syntenic groups, number of syntenic groups in the DCS region, number of *Tetraodon* chromosomes that alternate, and number of times the same combination of *Tetraodon* chromosomes occur in the human genome. See Supplementary Information for details.

Ancestral genome reconstruction

One category of DCS with the following definition encompassed most orthologues: "alternating series of i syntenic groups that belong to two ($i > = 2$) or three ($i > = 3$) *Tetraodon* chromosomes. The series may only be interrupted by groups from categories 'unassigned singletons' or 'background singletons'. A given combination of two or three *Tetraodon* chromosomes must appear at least twice in the human genome". These DCS blocks showed 12 recurring combinations of *Tetraodon* chromosomes, and were thus further classified in 12 groups labelled A to L. Each of the 12 groups, consisting of at least two DCS blocks with the same combination of alternating *Tetraodon* chromosomes, represents a proto-chromosome from the ancestral bony vertebrate (Osteichthyes). A model was then designed to account for the possible fates of chromosomes after duplication of the ancestral genome in the teleost lineage (Fig. 8). The model only deals with orthologous gene distribution between two genomes. It is simply based on the postulate that interchromosomal shuffling of genes within a genome increases with time, which is a measure to distinguish between ancient and recent events (for example, chromosome fusions or fissions). The two-dimensional distribution of 7,903 *Tetraodon*–human orthologues (Oxford Grid, Supplementary Fig. S12) was then confronted to the model and all 21 *Tetraodon* chromosomes could be grouped in pairs or triplets and assigned to a given type of event. See Supplementary Information for details.

Received 14 July; accepted 8 September 2004; doi:10.1038/nature03025.

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
3. Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
4. Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).

5. Hedges, S. B. The origin and evolution of model organisms. *Nature Rev. Genet.* **3**, 838–849 (2002).
6. Roest Crolius, H. *et al.* Human gene number estimate provided by genome wide analysis using *Tetraodon nigroviridis* genomic DNA. *Nature Genet.* **25**, 235–238 (2000).
7. Amores, A. *et al.* Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**, 1711–1714 (1998).
8. Robinson-Rechavi, M., Marchand, O., Escriva, H. & Laudet, V. An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Curr. Biol.* **11**, R458–R459 (2001).
9. Taylor, J. S., Braasch, L., Frickey, T., Meyer, A. & Van de Peer, Y. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res.* **13**, 382–390 (2003).
10. Batzoglou, S. *et al.* ARACHNE: a whole-genome shotgun assembler. *Genome Res.* **12**, 177–189 (2002).
11. Jaffe, D. B. *et al.* Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13**, 91–96 (2003).
12. Roest Crolius, H. *et al.* Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res.* **10**, 939–949 (2000).
13. Bouneau, L. *et al.* An active non-LTR retrotransposon with tandem structure in the compact genome of the pufferfish *Tetraodon nigroviridis*. *Genome Res.* **13**, 1686–1695 (2003).
14. Howe, K. L., Chothia, T. & Durbin, R. GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. *Genome Res.* **12**, 1418–1427 (2002).
15. Hatfield, D. L. *Selenium: Its Molecular Biology and Role in Human Health* (Kluwer, Dordrecht, 2001).
16. Boulay, J. L., O'Shea, J. J. & Paul, W. E. Molecular phylogeny within type I cytokines and their cognate receptors. *Immunity* **19**, 159–163 (2003).
17. Mulder, N. J. *et al.* InterPro: an integrated documentation resource for protein families, domains and functional sites. *Brief. Bioinform.* **3**, 225–235 (2002).
18. Dehal, P. *et al.* The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–2167 (2002).
19. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
20. Harris, M. A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32** (Database issue), D258–D261 (2004).
21. Torrents, D., Suyama, M., Zdobnov, E. & Bork, P. A genome-wide survey of human pseudogenes. *Genome Res.* **13**, 2559–2567 (2003).
22. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**, 57–86 (1986).
23. Gu, X. & Li, W. H. A general additive distance with time-reversibility and rate variation among nucleotide sites. *Proc. Natl Acad. Sci. USA* **93**, 4671–4676 (1996).
24. Holland, P. W. H. Introduction: gene duplication in development and evolution. *Semin. Cell Dev. Biol.* **10**, 515–516 (1999).
25. Martin, A. Is tetralogy true? Lack of support for the "one-to-four" rule. *Mol. Biol. Evol.* **18**, 89–93 (2001).
26. Wolfe, K. H. Yesterday's polyploids and the mystery of diploidization. *Nature Rev. Genet.* **2**, 333–341 (2001).
27. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
28. Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**, 304–307 (2004).
29. Prohaska, S. J. & Stadler, P. F. The duplication of the Hox gene clusters in teleost fishes. *Theor. Biosci.* **123**, 89–110 (2004).
30. Chiu, C. H. *et al.* Bichir HoxA cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Res.* **14**, 11–17 (2004).
31. Vandepoel, K., De Vos, W., Taylor, J. S., Meyer, A. & Van de Peer, Y. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl Acad. Sci. USA* **101**, 1638–1643 (2004).
32. Christoffels, A. *et al.* Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* **21**, 1146–1151 (2004).
33. Woods, I. G. *et al.* A comparative map of the zebrafish genome. *Genome Res.* **10**, 1903–1914 (2000).
34. Fischer, C. *et al.* Karyotype and chromosomal localization of characteristic tandem repeats in the pufferfish *Tetraodon nigroviridis*. *Cytogenet. Cell Genet.* **88**, 50–55 (2000).
35. Grutzner, F. *et al.* Classical and molecular cytogenetics of the pufferfish *Tetraodon nigroviridis*. *Chromosome Res.* **7**, 655–662 (1999).
36. Ohno, S., Wolf, U. & Atkin, N. B. Evolution from fish to mammals by gene duplication. *Heredity* **59**, 169–187 (1968).
37. Ojima, Y. in *Chromosomes in Evolution of Eukaryotic Groups* (eds Sharma, A. K. & Sharma, A.) 111–145 (CRC Press, Boca Raton, 1983).
38. Naruse, K. *et al.* A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res.* **14**, 820–828 (2004).
39. Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982).
40. Graves, J. A., Geciz, J. & Hameister, H. Evolution of the human X—a smart and sexy chromosome that controls speciation and development. *Cytogenet. Genome Res.* **99**, 141–145 (2002).
41. Richard, F., Lombard, M. & Dutrillaux, B. Reconstruction of the ancestral karyotype of eutherian mammals. *Chromosome Res.* **11**, 605–618 (2003).
42. The chromosome 21 mapping and sequencing consortium, The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
43. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
44. Collins, J. E. *et al.* Reevaluating human gene annotation: a second-generation analysis of chromosome 22. *Genome Res.* **13**, 27–36 (2003).
45. Heilig, R. *et al.* The DNA sequence and analysis of human chromosome 14. *Nature* **421**, 601–607 (2003).
46. Holland, P. W., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. Gene duplications and the

- origins of vertebrate development. *Development* (suppl.), 125–133 (1994).
47. Spring, J. Vertebrate evolution by interspecific hybridisation—are we polyploid? *FEBS Lett.* **400**, 2–8 (1997).
 48. Friedman, R. & Hughes, A. L. Pattern and timing of gene duplication in animal genomes. *Genome Res.* **11**, 1842–1847 (2001).
 49. Hughes, A. L., da Silva, J. & Friedman, R. Ancient genome duplications did not structure the human Hox-bearing chromosomes. *Genome Res.* **11**, 771–780 (2001).
 50. Thornton, J. W. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc. Natl Acad. Sci. USA* **98**, 5671–5676 (2001).
 51. McLysaght, A., Hokamp, K. & Wolfe, K. H. Extensive genomic duplication during early chordate evolution. *Nature Genet.* **31**, 200–204 (2002).
 52. Panopoulou, G. *et al.* New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res.* **13**, 1056–1066 (2003).
 53. Garcia-Fernandez, J. & Holland, P. W. Archetypal organization of the amphioxus Hox gene cluster. *Nature* **370**, 563–566 (1994).
 54. Artiguenave, F. *et al.* Genomic exploration of the hemiascomycetous yeasts: 2. Data generation and processing. *FEBS Lett.* **487**, 13–16 (2000).
 55. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
 56. Li, W. H., Wu, C. I. & Luo, C. C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150–174 (1985).

57. Pamilo, P. & Bianchi, N. O. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* **10**, 271–281 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by Consortium National de Recherche en Génomique. We thank T. Itami and S. Watabe for their gift of *Takifugu* blood samples; C. Nardon and M. Weiss for help with flow cytometry experiments; K. Howe for discussions regarding GAZE; R. Heilig for help with the annotation; the Centre Informatique National de l'Enseignement Supérieur for computer resources; and Gene-IT for assistance with the Biofacet software package.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.W. (jsbach@genoscope.cns.fr). The final assembly is available at EMBL/GenBank/DDBJ under accession number CAAE01000000. Full-length *Tetraodon* cDNAs have been deposited under accession numbers CR631133–CR735083; ultracontigs organized in chromosomes are available from <http://www.genoscope.org/tetraodon>.

A confirmation of the general relativistic prediction of the Lense–Thirring effect

I. Ciufolini¹ & E. C. Pavlis²

¹Dipartimento di Ingegneria dell'Innovazione, Università di Lecce and INFN Sezione di Lecce, Via Monteroni, 73100 Lecce, Italy

²Joint Center for Earth Systems Technology (JCET/UMBC), University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250, USA

An important early prediction of Einstein's general relativity^{1–3} was the advance of the perihelion of Mercury's orbit, whose measurement provided one of the classical tests of Einstein's theory⁴. The advance of the orbital point-of-closest-approach also applies to a binary pulsar system^{5,6} and to an Earth-orbiting satellite³. General relativity also predicts that the rotation of a body like Earth will drag the local inertial frames of reference around it^{3,7}, which will affect the orbit of a satellite⁸. This Lense–Thirring effect has hitherto not been detected with high accuracy⁹, but its detection with an error of about 1 per cent is the main goal of Gravity Probe B—an ongoing space mission using orbiting gyroscopes¹⁰. Here we report a measurement of the Lense–Thirring effect on two Earth satellites: it is 99 ± 5 per cent of the value predicted by general relativity; the uncertainty of this measurement includes all known random and systematic errors, but we allow for a total ± 10 per cent uncertainty to include underestimated and unknown sources of error.

Another general relativistic shift of a gyroscope has already been measured to an accuracy of about 0.35% using the Moon's orbit^{11,12}; the de Sitter or geodetic precession³. This shift arises from the effect of the gravitational field on the velocity of an orbiting gyroscope. But whereas the de Sitter effect is just due to the mass of a central non-rotating body, the Lense–Thirring effect is due to the rotation of a mass and displays the new general relativistic phenomenon that currents of mass generate additional space-time curvature³. This phenomenon has been called gravitomagnetism^{3,7} for its formal analogy with magnetism in electrodynamics. Though the Lense–Thirring effect may play a basic dynamical role in the accretion disk of a supermassive spinning black hole^{13,14} and in the alignment of jets in active galactic nuclei and quasars⁷, it is however extremely small on a satellite orbiting the Earth. For example, for a satellite with an orbital semimajor axis of about 12,000 km, the shift of its node (the intersection of the Earth's equatorial plane with the satellite's orbit, see Fig. 1) is only about 33 mas yr^{-1} (mas = millisecond of arc), that is, nearly 1.9 myr^{-1} . However, using the technique of laser-ranging with retro-reflectors to send back the short laser pulses, to date we are able to measure distances to a point on the Moon with a precision of a few centimetres, and distances to a small artificial satellite with a precision of a few millimetres.

In the present analysis we used the two laser-ranged satellites^{15,16} LAGEOS (NASA) and LAGEOS 2 (NASA-ASI). Their instantaneous position can be measured with a precision of a few millimetres and their orbits, with semimajor axes $a_{\text{LAGEOS}} \approx 12,270 \text{ km}$ and $a_{\text{LAGEOS2}} \approx 12,210 \text{ km}$, can be predicted, over 15-day periods, with a root-mean-square of the range residuals of a few centimetres. This uncertainty in the calculated orbits of LAGEOS and LAGEOS 2 is due to errors in modelling their orbital perturbations and, in particular, in modelling the deviations from spherical symmetry of the Earth's gravity field, mathematically described by a spherical harmonic expansion of the Earth's potential.

The current decade has been called the 'decade of geopotential

research', because we already have two unique missions in orbit, studying and mapping the gravitational field of the Earth and its temporal variations: DLR's CHAMP (Challenging Minisatellite Payload), launched in 2000, and NASA's GRACE (Gravity Recovery and Climate Experiment, see Fig. 1), launched on 17 March 2002, to be joined in orbit in mid-2006 by ESA's GOCE. GRACE consists of two identical spacecraft, very similar in design to CHAMP, orbiting in a polar orbit in tandem, some 200–250 km apart. Each of them carries a very precise accelerometer to measure non-gravitational forces, and they both range to each other via a K-band radar that produces a history of the inter-satellite distance variations to better than one micrometre. With non-gravitational forces adequately measured by the on-board accelerometers, the observed inter-satellite distance variations are used to determine the errors in the current models of the terrestrial gravitational field, thereby leading to improvements in the medium wavelengths of the model describing the field to a resolution of 200–250 km half wavelength. At the same time, tracking of both satellites by GPS (Global Positioning System) provides the information to improve the long-wavelength part of the model, including the zonal coefficients which describe the axially symmetric part of the Earth's potential. In the result presented here, we have used the recent Earth gravity model, EIGEN-GRACE02S, obtained by the GeoForschungsZentrum (GFZ) group using the GRACE satellites¹⁷. (The EIGEN-GRACE02S gravity field coefficients and their calibrated errors are available at http://op.gfz-potsdam.de/grace/index_GRACE.html; the GRACE mission is also described at <http://www.csr.utexas.edu>.) EIGEN-GRACE02S is accompanied by a set of calibrated error estimates (although still preliminary in nature), appropriate for the assessment of the final error budget of our result.

In our previous analyses⁹, we used the Earth gravity models JGM-3 and EGM96; owing to the limited accuracy associated with their low degree zonal terms, it was therefore necessary to use three observables: the node of LAGEOS and the node and perigee of LAGEOS 2. However, whereas the node is a very stable orbital

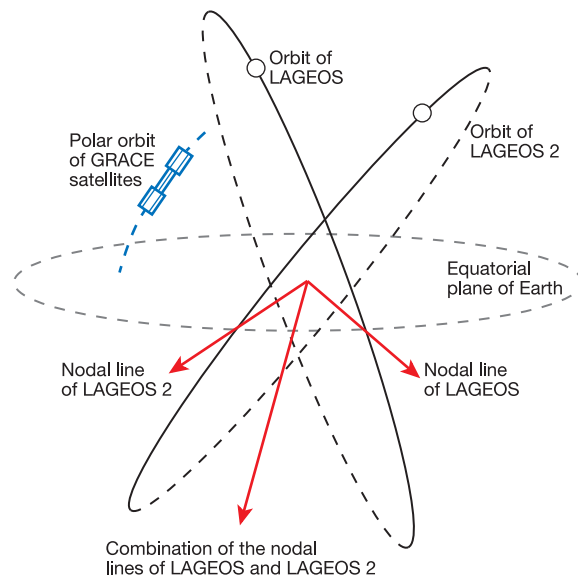


Figure 1 The 'orbital gyroscope' used to measure the Lense–Thirring effect. The 'gyroscope', indicated by the long red arrow, is the combination of the nodal longitudes of the LAGEOS satellites; it is not affected by the huge nodal rate of the LAGEOS satellites because of the Earth's quadrupole moment. This combination is described by equation (1); it is independent of the residual nodal rates due to the error in the Earth quadrupole moment. The blue drawing shows the orbital configuration of the GRACE satellites used to accurately determine the Earth's gravity field.

element under a number of non-gravitational perturbations, the perigee is affected by several perturbations difficult to model¹⁸, implying an accuracy not easy to be confidently assessed. Nevertheless, by analysing the uncertainties in the spherical harmonic coefficients of the recent Earth gravity model EIGEN-GRACE02S, we find that the only relevant uncertainty in the orbit of the LAGEOS satellites, comparing it with the magnitude of the Lense–Thirring effect, is the one, δJ_2 , in the Earth’s quadrupole moment, J_2 , which describes the Earth’s oblateness. In the EIGEN-GRACE02S model, the relative uncertainty $\delta J_2/J_2$ is about 10^{-7} . This uncertainty corresponds on the orbits of the LAGEOS satellites to a shift of the node larger than the Lense–Thirring effect. However, the orbital uncertainty due to all the other harmonics is only a few per cent of the general relativity shift. Therefore, in order to eliminate the orbital uncertainty due to δJ_2 and in order to solve for the Lense–Thirring effect, it is necessary and sufficient to use only two observables.

The two orbital observables that we analysed are the two nodes of the LAGEOS satellites (Fig. 1). After modelling all the orbital perturbations, apart from the Lense–Thirring effect, we are able to predict the LAGEOS satellites’ orbit with an error (root-mean-square of the residuals) of about 3 cm for a 15-day arc, corresponding to about half a millisecond of arc at the LAGEOS satellites’ altitude. The Lense–Thirring effect is, in contrast, 31 mas yr^{-1} on the LAGEOS node, and 31.5 mas yr^{-1} on the LAGEOS 2 node, as calculated by the Lense–Thirring formula: $\dot{\Omega} = 2J_{\oplus}/[a^3(1 - e^2)^{3/2}]$, where $\dot{\Omega}$ is the satellite’s rate of change of the nodal longitude, $J_{\oplus}$ is the Earth’s angular momentum, and a and e are the satellite’s semimajor axis and orbital eccentricity, respectively. The residual (calculated minus observed) nodal rate of the LAGEOS satellites, $\delta\dot{\Omega}^{\text{OBS}}$, is therefore: (residual nodal rate) = (nodal rate

from δJ_2 error) + (nodal rate from other δJ_{2n} errors) + (Lense–Thirring effect) + (other smaller modelling errors), where the δJ_{2n} are the errors in the Earth’s even zonal harmonic coefficients, J_{2n} , of degree $2n \geq 4$. We can then solve for the Lense–Thirring effect¹⁹ the system of the two observed residual nodal rates, $\delta\dot{\Omega}_I^{\text{OBS}}$, and simultaneously eliminate the error due to the δJ_2 uncertainty (see Methods). We accordingly combined the residuals as: $\delta\dot{\Omega}_I^{\text{OBS}} + k\delta\dot{\Omega}_{II}^{\text{OBS}}$, where the subscripts I and II indicate respectively LAGEOS and LAGEOS 2, and $k = 0.545$ is the ratio of the coefficients, K^2 , of J_2 in the nodal rate equations of LAGEOS and LAGEOS 2. In other words, $k = \frac{K_I^2}{K_{II}^2}$ is that coefficient that makes the combination $\dot{\Omega}_I + k\dot{\Omega}_{II}$ of the nodal rates of LAGEOS and LAGEOS 2 independent of any contribution of J_2 . Indeed, we have: $\dot{\Omega}_I(J_2) + k\dot{\Omega}_{II}(J_2) = K_I^2 J_2 + kK_{II}^2 J_2 = 0$, whereas by combining the nodal rates with the k factor we get for the Lense–Thirring contribution: $\dot{\Omega}_I^{\text{Lense-Thirring}} + k\dot{\Omega}_{II}^{\text{Lense-Thirring}} = (31 + k31.5) \text{ mas} \approx 48.2 \text{ mas}$. The maximum error in this combination of the residuals due to the δJ_{2n} is 4% of the Lense–Thirring effect (see Methods and Supplementary Information).

We analysed nearly eleven years of laser-ranging data, from January 1993 to December 2003, corresponding to about one million normal points, that is, to about 100 million laser ranging observations from more than 50 ILRS stations distributed all over the world²⁰. The total uncertainty of our measurement is, including systematic errors, $\pm 5\%$ of the Lense–Thirring effect and $\pm 10\%$ allowing for underestimated and unknown error sources (until more accurate Earth gravity field models are available, and until additional simulations are completed to exhaustively study the modelling of non-gravitational perturbations and measurement errors). For example, if we consider the time-independent gravitational error (root-sum-square) to be three times larger, we get a corresponding error of 9% and a total uncertainty of less than 10% (see Methods and Supplementary Information).

In Fig. 2 we show the observed residuals of the nodal longitudes of

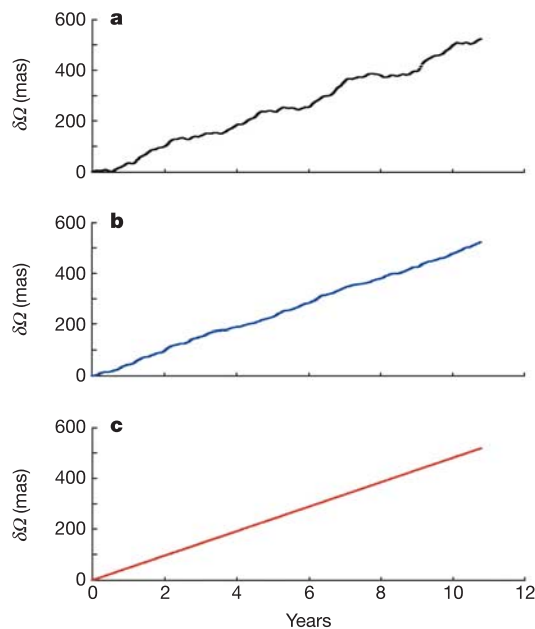


Figure 2 Observed orbital residuals of the LAGEOS satellites. The residual nodal longitudes of the LAGEOS satellites, $\delta\dot{\Omega}$, were combined according to equation (1). In black (a) is the raw, observed, residual nodal longitude of the LAGEOS satellites without removal of any signal, whereas in blue (b) is the observed residual nodal longitude after removal of six periodic signals. The best-fit line (13-parameter fit) through these observed residuals has a slope of 47.9 mas yr^{-1} . In red (c) is the theoretical Lense–Thirring prediction of Einstein’s general relativity for the combination (equation (1)) of the nodal longitudes of the LAGEOS satellites; its slope is 48.2 mas yr^{-1} .

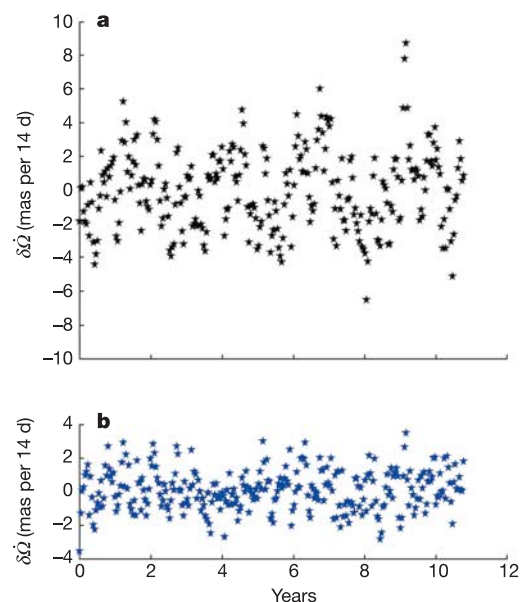


Figure 3 Post-fit orbital residuals of the LAGEOS satellites. These, 14-day, residuals of the nodal rates, $\delta\dot{\Omega}$, combined using equation (1), correspond to the case (a) of the fit of a secular trend only (black stars) and to the case (b) of a trend plus phase and amplitude of six periodic signals (blue stars) with periods of 1,044, 905, 281, 569, 111 and 284.5 days. We also fitted the residuals with a straight line plus the two LAGEOS nodal frequencies and plus ten signals. The maximum relative variation in the measured value of the Lense–Thirring effect in all the different fits was 2%.

the LAGEOS satellites, $\delta\Omega$, combined according to equation (1) in Methods, that is, we plot the residuals of the nodes of Fig. 1, combined with $k = 0.545$. The best fit line through the raw residuals in black (one-parameter fit) has a slope of 47.4 mas yr^{-1} ; the root-mean-square of these post-fit residuals is 15 mas. In blue are the residuals after removal of six main frequencies, corresponding to a 13-parameter fit with a secular trend plus phase and amplitude of six main signals (see Methods). In this case the secular trend is 47.9 mas yr^{-1} , but the root-mean-square of these post-fit residuals is only 6 mas. In red is the Lense–Thirring effect predicted by general relativity for the combination of the LAGEOS nodal longitudes, which amounts to 48.2 mas yr^{-1} . Therefore, corresponding to the 13-parameter fit (in blue) of Figs 2 and 3, the observed Lense–Thirring effect is 47.9 mas yr^{-1} , corresponding to 99% of the general relativistic prediction. In conclusion, this analysis confirms the general relativistic predictions of frame-dragging and the Lense–Thirring effect. □

Methods

Data analysis

The laser-ranging data were processed using NASA's orbital analysis and data reduction software GEODYN II²¹, over 15-day periods (with overlap of 1 day), following ref. 22, except for the use of the recent Earth static gravitational model EIGEN-GRACE02S. Solar radiation pressure, Earth albedo and anisotropic thermal thrust effects were modelled according to refs 23–26 using the LOSSAM-2004 model²⁷ (LAGEOS Spin Axis Model) of the satellites' spin axis, and the Earth tides using the GOT99.2 (Goddard/Grenoble Ocean Tide) model²⁸. Lunar, solar and planetary perturbations (JPL ephemerides DE-403) were included in the equations of motion and Earth's rotation was modelled from very long baseline interferometry and GPS. The Lense–Thirring effect was set equal to zero. For every 14-day period we obtained one node residual for LAGEOS and one for LAGEOS 2, which were then combined according to equation (1):

$$\delta\Omega_{\text{I}}^{\text{OBS}} + k\delta\Omega_{\text{II}}^{\text{OBS}} = \delta\Omega_{\text{I}}^{\text{Lense-Thirring}} + k\delta\Omega_{\text{II}}^{\text{Lense-Thirring}} + \sum_{2n \geq 4} (K_1^{2n} |\delta J_{2n}| + k K_2^{2n} |\delta J_{2n}|) \quad (1)$$

where $k = 0.545$ and K^{2n} , well determined functions of the orbital elements, are the coefficients in the nodal rate equations of the even zonal harmonics (even degree and zero order), J_{2n} , of the Earth's gravity field.

Error assessment

The dominant error in modelling the nodal drift of the LAGEOS satellites is due to the errors, δJ_{2n} , in the static part of the J_{2n} . We assessed this error by using the EIGEN-GRACE02S model, with its δJ_{2n} calibrated (that is, including systematic errors) uncertainties, and using our combination, equation (1), to eliminate the largest uncertainty, δJ_2 , in the Earth's quadrupole moment. By taking the square root of the sum of the squared errors due to the δJ_{2n} in equation (1), we found a relative error of about 3% of the Lense–Thirring effect on the LAGEOS satellites. However, to get an upper bound to this error (since the covariance matrix was not available to us) we simply added the absolute values of the errors due to the δJ_{2n} in equation (1); we thus obtained an error estimate of about 4% of the Lense–Thirring effect.

To assess the error in modelling the nodal drifts due to all the other orbital perturbations, we used two different methods: (1) an *a priori* detailed analysis of the uncertainties in the perturbations affecting the LAGEOS satellites, based on the extensive scientific literature on this subject^{23–27,29–31}, and (2) an *a posteriori* analysis of the orbital residuals. Both methods produced the same result. Using the *a priori* error analysis, for the gravitational perturbations (solar and lunar Earth tides, secular trends in the even zonal harmonics of the Earth's field and other periodic variations in the Earth's harmonics), we obtained, over a period of 11 yr, a total relative uncertainty of about 2% of the Lense–Thirring effect³⁰. For the non-gravitational perturbations (atmospheric drag, solar radiation pressure, Earth albedo, anisotropic thermal radiation from the satellites, both from solar radiation and from the Earth's infrared radiation heating) we got, over a period of 11 yr, a total relative uncertainty of about 2% of the Lense–Thirring effect³⁰. We also included an error of about 2% due to stochastic errors, such as seasonal variations in the Earth's gravity field, and observation biases.

Using the orbital residuals and considering that the main periods of the non-gravitational perturbations and of the tidal effects are well determined^{23–31}, we performed a number of different fits of the residuals. We fitted the residuals with a secular trend only, or with a trend plus a different number of the main periodic terms (fitting both phase and amplitude). These terms correspond to various linear combinations of the LAGEOS satellites' nodal periods (1,044 and 569 days), the Earth's revolution (365.25 days) and the Moon's nodal period (18.61 yr). In addition, we also performed a Fourier analysis of the residuals and removed the main identified frequencies in our fit. The maximum variation we found in all these different fits was at most a 2% change in the measured value of the Lense–Thirring effect, confirming that the error in the modelling of the periodic perturbations may in the worst case affect our Lense–Thirring determination at a level of about 2% only.

Total uncertainty

Finally, considering all the error sources described above and in agreement with previous error analysis of the LAGEOS III-LARES experiment^{29–31} (see also Supplementary Discussion), we obtained a total root-sum-square error of $\pm 5\%$ of the Lense–Thirring effect. This uncertainty refers to all the known errors, however, allowing for unknown and unmodelled error sources we assume a $\pm 10\%$ uncertainty in our measurement. For example, if we double the maximum time-independent gravitational error and the non-gravitational and time-dependent gravitational errors, we get respectively 8%, 4% and 4% errors, and thus a total uncertainty of 10% of the Lense–Thirring effect.

Received 2 June; accepted 10 September 2004; doi:10.1038/nature03007.

- Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation* (Freeman, San Francisco, 1973).
- Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).
- Ciufolini, I. & Wheeler, J. A. *Gravitation and Inertia* (Princeton Univ. Press, Princeton, New Jersey, 1995).
- Will, C. M. *Theory and Experiment in Gravitational Physics* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 1993); The confrontation between general relativity and experiment. (<http://www.livingreviews.org/Articles/Volume4/2001-4will>) (2001).
- Hulse, R. A. & Taylor, J. H. Discovery of a pulsar in a binary system. *Astrophys. J.* **195**, L51–L53 (1975).
- Nordtvedt, K. Existence of the gravitomagnetic interaction. *Int. J. Theor. Phys.* **27**, 1395–1404 (1988).
- Thorne, K. S., Price, R. H. & Macdonald, D. A. *The Membrane Paradigm* (Yale Univ. Press, New Haven, 1986).
- Lense, J. & Thirring, H. Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie. *Phys. Z.* **19**, 156–163 (1918).
- Ciufolini, I. in *A Relativistic Spacetime Odyssey, Proc. 25th Johns Hopkins Workshop on Current Problems in Particle Theory* (eds Ciufolini, I., Dominici, D. & Lusanna, L.) 99–139 (World Scientific, Singapore, 2003).
- Fitch, V. L. *et al. Review of Gravity Probe B* (National Academic Press, Washington DC, 1995).
- Williams, J. G., Boggs, D. H., Dickey, J. O. & Folkner, W. M. in *Proc. 9th Marcel Grossmann Meeting* (eds Gurzadyan, V. G., Jantzen, R. T. & Ruffini, R.) 1797–1798 (World Scientific, Singapore, 2002).
- Shapiro, I. I., Reasenberg, R. D., Chandler, J. F. & Babcock, R. W. Measurement of the de Sitter precession of the Moon. *Phys. Rev. Lett.* **61**, 2643–2646 (1988).
- Rees, M. J. in *Black Holes in Binaries and Galactic Nuclei, Proc. ESO Workshop* (eds Kaper, L., van den Heuvel, E. P. J. & Woudt, P. A.) 351–363 (Springer, Heidelberg, 2001).
- Genzel, R. *et al.* Near-infrared flares from accreting gas around the supermassive black hole at the Galactic Centre. *Nature* **425**, 934–937 (2003).
- LAGEOS scientific results. *J. Geophys. Res.* **B 90**, 9215–9438 (1985).
- Bender, P. & Goad, C. C. in *Proc. 2nd Int. Symp. on the Use of Artificial Satellites for Geodesy and Geodynamics Vol. II* (eds Veis, G. & Livieratos, E.) 145 (National Technical University of Athens, Athens, 1979).
- Reigber, C. *et al.* An Earth gravity field model complete to degree and order 150 from GRACE: EIGEN-GRACE02S. *J. Geodyn.* (in the press).
- Bertotti, B., Farinella, P. & Vokrouhlický, D. *Physics of the Earth and the Solar System* 2nd edn (Kluwer Academic, Dordrecht, 2003).
- Ciufolini, I. Measurement of the Lense–Thirring drag on high-altitude laser-ranged artificial satellites. *Phys. Rev. Lett.* **56**, 278–281 (1986).
- Noomen, R., Klosko, S., Noll, C. & Pearlman, M. (eds) Toward millimeter accuracy. *Proc. 13th Int. Laser Ranging Workshop* (NASA CP 2003–212248, NASA Goddard, Greenbelt, MD, 2003).
- Pavlis, D. E. *et al.* *GEODYN Systems Description Vol. 3* (NASA GSFC, Greenbelt, MD, 1998).
- McCarthy, D. *The 1996 IERS Conventions* (Observatoire de Paris, Paris, 1996).
- Rubincam, D. P. Yarkovsky thermal drag on LAGEOS. *J. Geophys. Res.* **B 93**, 13805–13810 (1988).
- Rubincam, D. P. Drag on the LAGEOS satellite. *J. Geophys. Res.* **B 95**, 4881–4886 (1990).
- Rubincam, D. P. & Malla, A. Terrestrial atmospheric effect on satellite eclipses with application to the acceleration of LAGEOS. *J. Geophys. Res.* **B 100**, 20285–20290 (1995).
- Martin, C. F. & Rubincam, D. P. Effects of Earth albedo on the LAGEOS I satellite. *J. Geophys. Res.* **B 101**, 3215–3226 (1996).
- Andrés, J. I. *et al.* Spin axis behavior of the LAGEOS satellites. *J. Geophys. Res.* **109**, B06403 (2004).
- Ray, R. D. *A Global Ocean Tide Model from Topex/Poseidon Altimetry: GOT99.2* (NASA Tech. Memo. 209478, NASA GSFC, Greenbelt, MD, 1999).
- Ciufolini, I. A comprehensive introduction to the LAGEOS gravitomagnetic experiment. *Int. J. Mod. Phys. A* **4**, 3083–3145 (1989).
- Ciufolini, I. *et al.* *WEBER-SAT/LARES Study for INFN* (Università di Lecce, Lecce, 2004).
- Ries, J. C., Eanes, R. J., Watkins, M. M. & Tapley, B. in *LARES Phase A Study for ASI* (ed. Ciufolini, I.) (Italian Space Agency, Rome, 1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the International Laser Ranging Service (ILRS) for making available for our use the laser ranging data collected by their network, through their data service at CDDIS, NASA/Goddard, USA, and the GeoForschungsZentrum, Potsdam, Germany. E.C.P. acknowledges support from NASA. I.C. thanks R. Matzner and N. Cabibbo for comments. We thank R. Peron for reading the manuscript and, together with G. Chionchio, D. Lucchesi, D. Pavlis and F. Ricci, for computer support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.C. (ignazio.ciufolini@unile.it).

Poisson's ratio and the fragility of glass-forming liquids

V. N. Novikov^{1,2} & A. P. Sokolov¹

¹Department of Polymer Science, The University of Akron, Akron, Ohio 44325-3909, USA

²IA&E, Russian Academy of Sciences, Novosibirsk 630090, Russia

The nature of the transformation by which a supercooled liquid 'freezes' to a glass—the glass transition—is a central issue in condensed matter physics^{1–3} but also affects many other fields, including biology⁴. Substantial progress has been made in understanding this phenomenon over the past two decades, yet many key questions remain. In particular, the factors that control the temperature-dependent relaxation and viscous properties of the liquid phase as the glass transition is approached (that is, whether the glass-forming liquid is 'fragile' or 'strong'^{5–7}) remain unclear. Here we show that the fragility of a glass-forming liquid is intimately linked to a very basic property of the corresponding glass phase: the relative strength of shear and bulk moduli, or Poisson's ratio.

Fragility of liquids is defined as the apparent activation energy of shear viscosity η or structural relaxation time τ_α at the glass transition temperature T_g , normalized to T_g (ref. 7):

$$m = \frac{\partial \log \eta(T)}{\partial (T_g/T)} \bigg|_{T=T_g} \quad (1)$$

It characterizes the steepness of the slope of $\log \eta$ dependence on T_g/T near T_g (Fig. 1). A stronger deviation from Arrhenius behaviour corresponds to a more fragile system. One of the puzzles that remains unsolved is the correlation of the fast terahertz dynamics and fragility^{8,9}: one can predict fragility, that is, temperature variations of τ_α or η in a supercooled liquid at $T \approx T_g$, from analysis of vibrational spectra at terahertz frequencies deep in the glassy state ($T \ll T_g$)^{8,9}. Interest in this puzzle was recently stimulated by ref. 9 (see also refs 10 and 11). Our present goal is to gain better insight into fragility on the microscopic scale and to suggest a consistent explanation for these correlations.

It is well understood that the glass transition is a failure of the material to support shear stress on the typical laboratory timescale of, say, a few minutes. So, shear modulus G is apparently an important parameter. To be more precise, the modulus is frequency-dependent and one needs to consider G at frequencies much higher than the inverse structural relaxation time, that is, the so-called instantaneous shear modulus G_∞ . In an isotropic solid like glass there is only one different independent elastic constant that makes it possible to get the dimensionless parameter proportional to G_∞ , namely, the instantaneous bulk modulus K_∞ . The ratio of instantaneous shear and bulk moduli can be estimated from the ratio of longitudinal ($v_l^2 = M_\infty/\rho$) and transverse ($v_t^2 = G_\infty/\rho$) sound velocities in the glassy state, for example, $K_\infty/G_\infty \propto (v_l/v_t)^2 - 4/3$ (here ρ is the mass density and $M_\infty = K_\infty + (4/3)G_\infty$ is the instantaneous longitudinal modulus). We note that the high-frequency sound velocity in glasses is connected with the adiabatic bulk modulus, although the difference between the isothermal and adiabatic bulk moduli in the glassy state is very small. Analysis of a large number of glasses, including covalent and hydrogen-bonded, van-der-Waals and ionic glasses, shows a correlation between v_l/v_t and m (Fig. 2): the more weakly the system resists the shear stress in comparison with the bulk one in the glassy state (higher v_l/v_t), the more fragile its behaviour appears in the melt. We emphasize that the data presented in Fig. 2 includes all the systems we were able to find in the literature, excluding polymers. It is known that the

fragility of some polymers strongly depends on molecular weight, M_w (ref. 12). For example, the fragility of polystyrene varies from 60 at low M_w up to 150 at high M_w (ref. 12), and only low- M_w polystyrene is in agreement with the correlation of Fig. 2. This sharp increase in fragility with M_w might be polymer-specific; for this reason, all polymers were excluded from consideration, although low fragility polymers like polyisobutylene and polybutadiene agree well with the correlation of Fig. 2.

The inset in Fig. 2 shows the correlation of fragility m with K_∞/G_∞ of the respective glass; it can be well described by the relationship:

$$m - 17 = 29(K_\infty/G_\infty - 1) \text{ or } m = 29(K_\infty/G_\infty - 0.41) \quad (2)$$

where 17 is the lowest value expected for m and K_∞/G_∞ in glasses is not expected to be lower than 1 (for strong glasses like silica and BeF₂, $K_\infty/G_\infty \approx 1.1$). We note that correlation (equation (2)) means also that fragility of a liquid is fully determined just by the Poisson ratio of its glass.

How can one explain the correlation between the fragility of a liquid and the ratio of the instantaneous elastic moduli of a glass (Fig. 2)? The answer might be buried deep in the liquid state because the structure and properties of a glass are essentially the structure and properties of a 'frozen' liquid. Owing to the construction of the fragility plot, all viscosity or relaxation time curves intersect at two points: (1) at T_g , $\log \eta \equiv \log \eta_g = 13$ (where η is in poise), or $\log \tau_\alpha \approx 3$ (where τ is in seconds); and (2) at very high temperatures, $T_g/T \rightarrow 0$, where all liquids have $\log \eta \equiv \log \eta_0 \approx -4$ (ref. 13) or $\log \tau_\alpha \approx \log \tau_0 \approx -14$. This means that if a liquid has a steeper slope of $\log \eta$ near T_g , it inevitably has a smaller slope of $\log \eta$ at high temperature. At high temperatures, relaxation in most of the liquids shows Arrhenius temperature dependence: $\eta = \eta_0 \exp(E/T)$. Thus the high-temperature slope of $\log \eta$ in Fig. 1, E/T_g , can also be a measure of fragility. Experimental data show that E/T_g indeed correlates with fragility m (Fig. 3):

$$\frac{E}{T_g} = \frac{(19.2)^2 \ln 10}{m} \quad (3)$$

We note that a similar relationship with a slightly different coefficient $(17)^2 \ln 10$ follows from the Vogel–Fulcher–Tamman (VFT) equation, $\eta = \eta_0 \exp B/(T - T_0)$, if one assumes that the VFT equation is valid over the entire temperature range. The quantitative disagreement can be related to the well-known fact that a single VFT equation cannot accurately describe η at all temperatures¹⁴.

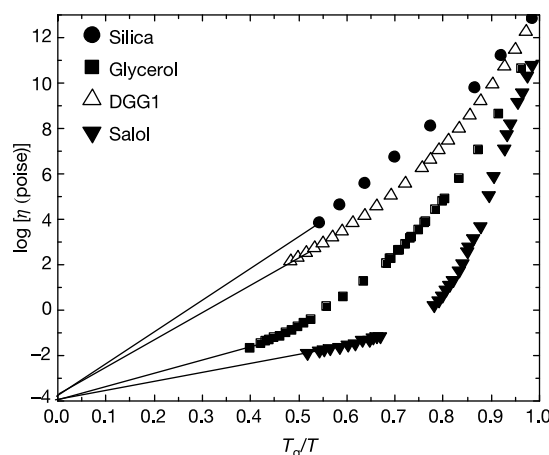


Figure 1 An example of a fragility plot. Data is from refs 2, 27 and 28. DGG1 is a soda-lime silica glass; see Supplementary Table 1. The solid lines show the expected high-temperature slopes E/T_g .

This connection between m and T_g/E helps to explain the correlation between m and the ratio of instantaneous elastic moduli in the glassy state, equation (2). Various authors^{6,15,16} suggest that the activation energy of viscosity is proportional to the instantaneous shear modulus G_∞ , that is, $E \propto G_\infty V_c$, where V_c is some volume that does not show significant temperature variations at high T . On the other hand, there are empirical correlations between elastic moduli of a glass and T_g (for example, refs 6, 17 and 18). However, it is not obvious which combination of shear and bulk moduli is important because these correlations were usually considered within a class of materials with similar chemical structure, and thus with a similar Poisson's ratio. We note that the free volume model for the glass transition¹⁸ predicts $T_g \propto K_\infty$. Analysis of experimental data reveals correlation of T_g/E to the same parameter ($K_\infty/G_\infty - 0.41$) as in equation (2):

$$T_g/E \approx 0.037(K_\infty/G_\infty - 0.41) \propto m \quad (4)$$

Thus, the correlation found (equations (2) or (4) and Fig. 2) emphasizes a very simple rule: the better the glass can resist shear deformation rather than dilatation, the stronger (that is, more Arrhenius-like) the behaviour that is exhibited during its structural relaxation.

On a timescale longer than that of structural relaxation, $t \gg \tau_\alpha$, the shear modulus (for non-polymeric systems) relaxes to zero while the longitudinal modulus relaxes to some finite value M_0 . This leads to a difference between instantaneous (v_∞) and zero-frequency (v_0) longitudinal sound velocities in glasses. The difference provides an estimate of the relaxation strength of the α -process, that is, the so-called non-ergodicity parameter, $f_0 = 1 - v_0^2/v_\infty^2$. A sizeable part of the decrease of the longitudinal sound velocity at $\omega\tau_\alpha \ll 1$ is due to relaxation of the shear modulus. Because $M = K + (4/3)G$, a crude approximation then gives:

$$f_0 \approx (4/3)v_t^2/v_l^2 = G_\infty/[(3/4)K_\infty + G_\infty] \quad (5)$$

Thus, the ratio G_∞/K_∞ also determines the amplitude of the non-ergodicity parameter.

Equations (5) and (4) lead us directly to an explanation of the

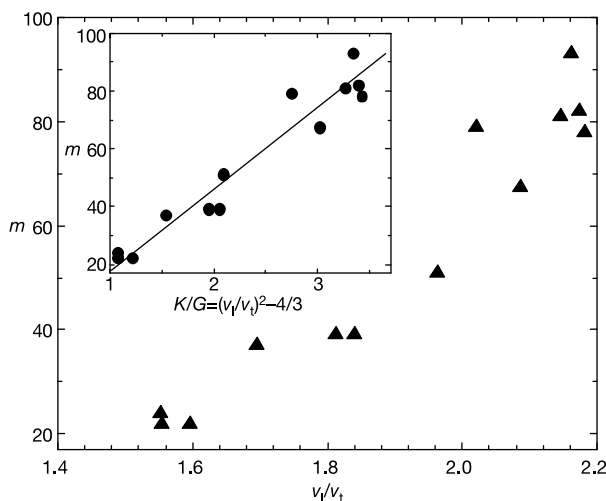


Figure 2 Correlation between fragility and the ratio of longitudinal and transversal sound velocities found in the glassy state. Inset, correlation of fragility with the ratio of the bulk and shear moduli in the glassy state. The straight line shows the relationship from equation (2). Materials (in descending order of fragility): CKN ($\text{Ca}_2(\text{NO}_3)_2\text{-KNO}_3$), m-TCP (m-tricresyl phosphate), OTP (o-therphenyl), Se, m-toluidine, salol, glycerol, B_2O_3 , As_2S_3 , $20(\text{Na}_2\text{O})80(\text{SiO}_2)$, SiO_2 , GeO_2 , BeF_2 . In those cases where more than one value of fragility is known, we used the average value. Data from the literature for fragility, v_t and v_l and the respective references are given in Supplementary Table 1.

recent puzzling observation reported in ref. 9. The authors found that the parameter α , defined as $\alpha = R_{\text{LP}}^{-1}(T)T_g/T$, measured using inelastic X-ray scattering in a few glasses, correlates with their fragility, $\alpha \propto m$. (Here $R_{\text{LP}}(T) = I_c/2I_B$ is the Landau–Placzek ratio, and $2I_B$ and I_c are the integrated intensities of the combined Brillouin doublet and the central line of the dynamic structure factor $S(q, \omega)$, respectively.) It is known^{10,19} that in the case of plane-wave phonons $R_{\text{LP}}(T) = \kappa_T M_{\text{Br}} - 1$, where κ_T is the isothermal compressibility, and M_{Br} is the adiabatic longitudinal modulus at the Brillouin line frequency. Using the relationships $v_0^2 \approx 1/\rho\kappa_T$ and $v_\infty^2 \approx M_\infty/\rho$, α can be expressed via the non-ergodicity parameter at T_g :

$$\alpha = (1 - f_0)/f_0 \quad (6)$$

$(1 - f_0)/f_0$ indeed correlates well with fragility¹⁰, as also does the non-ergodicity parameter f_0 . Substituting equation (5) into equation (6) gives $\alpha \propto K_\infty/G_\infty$. This provides a clear microscopic interpretation of the correlation between α and m reported in ref. 9: G_∞/K_∞ controls both the relative amplitude of the structural relaxation in a glass and the fragility of a supercooled liquid.

$f_0(T_g)$ also determines the amplitude of fluctuations frozen at T_g . That leads us to another unsolved problem—correlation between fragility and the amplitude of the boson peak²⁰. All disordered materials have an excess density of vibrational states, $g(\nu)$, over the Debye density of states, $g_{\text{Deb}}(\nu)$, in the terahertz frequency range^{20–24}, so that the ratio $g(\nu)/g_{\text{Deb}}(\nu)$ exhibits the so-called boson peak with amplitude A_{bp} . According to refs 20–25, A_{bp} is related to the amplitude of frozen structure fluctuations. So, one would expect higher A_{bp} in materials with higher $f_0(T_g)$ and thus A_{bp} should correlate with v_t/v_l , and, consequently, with fragility. Comparison of $(v_t/v_l)^2$ and A_{bp} in a few glasses indeed shows very good correlation (Supplementary Fig. 1). Thus, the correlation of the boson peak amplitude to fragility observed in ref. 20 seems to be related to the same role of the non-ergodicity parameter. In other words, the capability of a structure to resist shear rather than bulk deformation also determines the boson peak amplitude through the amplitude of frozen fluctuations, $f_0(T_g)$.

Recently it was shown that higher anharmonicity of a glass-former corresponds to higher fragility^{8,11,26}. To understand this, we note that anharmonicity always leads to an increase of the fast relaxation in glasses²⁶, and thus, to a decrease of f_0 . Because, as we showed above, f_0 correlates with fragility, the same should be valid

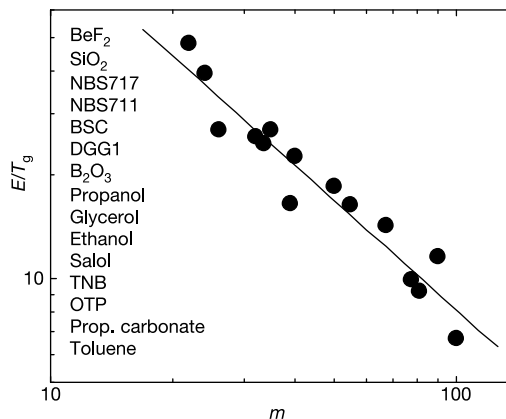


Figure 3 Correlation between fragility and the ratio of the high-temperature activation energy to the glass transition temperature. Materials are listed in ascending fragility order, where NBS717 is borosilicate glass, NBS711 is lead silica glass, BSC is borosilicate crown glass, TNB is 1,3,5-tri- α -naphthyl benzene and prop. is propylene. Data from the literature for fragility, the high-temperature slope of the fragility plot, E/T_g , and the respective references are given in Supplementary Table 1.

for the anharmonicity. A similar explanation can be applied to the correlation between the intensity of the fast relaxation normalized to the boson peak and the fragility found in ref. 8. The amplitude of the fast relaxation is related to $1 - f_0$, so the ratio of the fast relaxation to the boson peak should be proportional to $(1 - f_0)/f_0$, and correlate linearly with α and m .

Thus the ratio of instantaneous shear to bulk modulus in glasses, or, alternatively, the Poisson ratio, appears to be an important parameter that controls the fragility of liquids. This ratio also determines the non-ergodicity parameter f_0 , that is, how limited is the motion of a molecular unit in a cage formed by its neighbours. The same non-ergodicity parameter controls behaviour of the fast dynamics and that explains the correlation between fragility and fast dynamics in glass-forming systems. □

Received 1 June; accepted 17 August 2004; doi:10.1038/nature02947.

1. Anderson, P. W. Through a glass lightly. *Science* **267**, 1615 (1995).
2. Angell, C. A. Formation of glasses from liquids and biopolymers. *Science* **267**, 1924–1935 (1995).
3. Debenedetti, P. G. & Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **410**, 259–267 (2001).
4. Fox, K. C. Putting proteins under glass. *Science* **267**, 1922–1923 (1995).
5. Nemilov, S. V. The kinetics of elementary processes in the condensed state. II. Shear relaxation and the equation of state for solids. *Zh. Fiz. Khim.* **42**, 1391–1396 (1968); *Russ. J. Phys. Chem.* **42**, 726–731 (1968).
6. Nemilov, S. V. *Thermodynamic and kinetic aspects of the vitreous state* (CRC Press, Boca Raton, 1995).
7. Angell, C. A. Spectroscopy simulation and scattering, and the medium range order problem in glass. *J. Non-Cryst. Solids* **73**, 1–17 (1985).
8. Sokolov, A. P., Rössler, E., Kisliuk, A. & Quitmann, D. Dynamics of strong and fragile glassformers: differences and correlation with low-temperature properties. *Phys. Rev. Lett.* **71**, 2062–2065 (1993).
9. Scopigno, T., Ruocco, G., Sette, F. & Monaco, G. Is the fragility of a liquid embedded in the properties of its glass? *Science* **302**, 849–852 (2003).
10. Buchenau, U. & Wischniewski, A. Fragility and compressibility at the glass transition. Preprint at <http://arXiv.org/cond-mat/0401088> (2004).
11. Bordat, P., Affouard, F., Descamps, M. & Ngai, K. L. Does the interaction potential determine both the fragility of a liquid and the vibrational properties of its glassy state? Preprint at <http://arXiv.org/cond-mat/0401117> (2004).
12. Roland, C. M. & Casalini, R. Temperature dependence of local segmental motion in polystyrene and its variation with molecular weight. *J. Chem. Phys.* **119**, 1838–1842 (2003).
13. Barrer, R. M. The viscosity of pure liquids. I. Non-polymerised fluids. *Trans. Faraday Soc.* **39**, 48–59 (1943).
14. Stickel, F., Fischer, E. W. & Richert, R. Dynamics of glass-forming liquids. II. Detailed comparison of dielectric relaxation, dc-conductivity, and viscosity data. *J. Chem. Phys.* **104**, 2043–2055 (1996).
15. Tobolsky, A. V., Powell, R. E. & Eyring, H. In *Frontiers in Chemistry* (eds Burk, R. E. & Grummit, O.) Vol. 1, 125 (Interscience, New York, 1943).
16. Dyre, J. C. & Olsen, N. B. Landscape equivalent of the shoving model. *Phys. Rev. E* **69**, 042501–042504 (2004).
17. Wang, W. H., Wen, P., Zhao, D. Q., Pan, M. X. & Wang, R. J. Relationship between glass transition temperature and Debye temperature in bulk metallic glasses. *J. Mater. Res.* **18**, 2747–2751 (2003).
18. Sanditov, D. S., Sangadiev, S. Sh. & Kozlov, G. V. On the correlation between elastic modulus of vitreous solids and glass transition temperature of melts. *Glass Phys. Chem.* **24**, 539–547 (1998).
19. Fabelinsky, I. L. *Molecular Scattering of Light* Ch. VII (Plenum, New York, 1968).
20. Sokolov, A. P. et al. Low-temperature anomalies in strong and fragile glass formers. *Phys. Rev. Lett.* **78**, 2405–2408 (1997).
21. Courtens, E., Foret, M., Hehlen, B. & Vacher, R. The vibrational modes of glasses. *Solid State Commun.* **117**, 187–200 (2001).
22. Grigera, T. S., Martin-Mayor, V., Parisi, G. & Verrocchio, P. Phonon interpretation of the 'boson peak' in supercooled liquids. *Nature* **422**, 289–292 (2003).
23. Malinovsky, V. K. & Sokolov, A. P. The nature of boson peak in Raman spectra in glasses. *Solid State. Comm.* **57**, 757–761 (1986).
24. Duval, E., Boukenter, A. & Achibat, T. Vibrational dynamics and the structure of glasses. *J. Phys. Condens. Matter* **2**, 10227–10234 (1990).
25. Soltwisch, M., Quitmann, D. & Ruocco, G. On the connection between low frequency vibrational and relaxational motion in glasses. *J. Non-Cryst. Solids* **203**, 12–18 (1996).
26. Novikov, V. N. Vibration anharmonicity and fast relaxation in the region of glass transition. *Phys. Rev. B* **58**, 8367–8378 (1998).
27. Martinez, L.-M. & Angell, C. A. A thermodynamic connection to the fragility of glass-forming liquids. *Nature* **410**, 663–667 (2001).
28. Rössler, E., Hess, K.-U. & Novikov, V. N. Universal representation of viscosity in glass forming liquids. *J. Non-Cryst. Solids* **223**, 207–222 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by NSF and REFI grants.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.P.S. (alexsei@uakron.edu).

Low-voltage organic transistors with an amorphous molecular gate dielectric

Marcus Halik¹, Hagen Klauk¹, Ute Zschieschang¹, Günter Schmid¹, Christine Dehm¹, Markus Schütz², Steffen Maisch², Franz Effenberger², Markus Brunnbauer³ & Francesco Stellacci³

¹Infineon Technologies AG, New Memory Platforms, Materials and Technology, Paul-Gossen-Straße 100, 91052 Erlangen, Germany

²University Stuttgart, Department of Chemistry, Pfaffenwaldring 55, 70569 Stuttgart, Germany

³Massachusetts Institute of Technology, Department of Materials Science and Engineering, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

Organic thin film transistors (TFTs) are of interest for a variety of large-area electronic applications, such as displays^{1–3}, sensors^{4,5} and electronic barcodes^{6–8}. One of the key problems with existing organic TFTs is their large operating voltage, which often exceeds 20 V. This is due to poor capacitive coupling through relatively thick gate dielectric layers: these dielectrics are usually either inorganic oxides or nitrides^{2–8}, or insulating polymers⁹, and are often thicker than 100 nm to minimize gate leakage currents. Here we demonstrate a manufacturing process for TFTs with a 2.5-nm-thick molecular self-assembled monolayer (SAM) gate dielectric and a high-mobility organic semiconductor (pentacene). These TFTs operate with supply voltages of less than 2 V, yet have gate currents that are lower than those of advanced silicon field-effect transistors with SiO₂ dielectrics. These results should therefore increase the prospects of using organic TFTs in low-power applications (such as portable devices). Moreover, molecular SAMs may even be of interest for advanced silicon transistors where the continued reduction in dielectric thickness leads to ever greater gate leakage and power dissipation.

The use of SAM gate dielectrics for organic TFTs was pioneered by the Vuillaume group¹⁰. They investigated the mechanism of carrier tunnelling through SAMs and showed that current leakage through densely packed organic molecular monolayers with highly ordered aliphatic chains can be remarkably low, despite their thickness of only a few nanometres¹¹. To prepare organic TFTs (and even silicon metal–insulator–semiconductor field-effect transistors, MISFETs) with SAM dielectrics, the Vuillaume group relied on carboxyl-terminated *n*-alkyltrichlorosilanes^{10,12}. Using α -sexithiophene as the organic semiconductor, they demonstrated organic TFTs operating at 2 V, with a subthreshold swing of 350 mV per decade, a field-effect mobility of $3.6 \times 10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, an on/off current ratio of 10^4 , and a gate current density of $10^{-6} \text{ A cm}^{-2}$ (ref. 10).

Given the small thickness of the molecular monolayers (about 2.5 nm), a gate current density of $10^{-6} \text{ A cm}^{-2}$ may seem reasonably low. However, the current density through a molecular monolayer can be less than $10^{-8} \text{ A cm}^{-2}$ (at 1 V) if no semiconductor is deposited¹¹. One possible explanation for the increase in gate current upon deposition of the organic semiconductor is the partial penetration of the SAM by the semiconductor molecules. This penetration is likely to reduce the dielectric thickness and create low-resistance current paths through the SAM. Such interaction between the SAM dielectric and the organic semiconductor is undesirable not only because it leads to increased gate leakage, but also because high-mobility carrier transport in the semiconductor depends critically on a well-defined semiconductor–dielectric interface.

To create dense self-assembled monolayers with sufficient robustness against molecular penetration, we have used a specifically

synthesized alkyltrichlorosilane with an aromatic endgroup¹³. The chemical structure of the molecule, (18-phenoxyoctadecyl)trichlorosilane (PhO-OTS), is shown in Fig. 1. The SAMs are created in a one-step process without the need for chemical conversion^{10,12,14}. Upon self-assembly on a natively oxidized silicon surface, the π - π interaction between the phenoxy endgroups of adjacent molecules creates an intermolecular top-link, leading to a more closely packed surface compared with monolayers with linear endgroups. As a result, SAMs prepared using PhO-OTS not only show very low leakage currents, but allow the use of pentacene as the organic semiconductor. Of all organic semiconductors that have been used to prepare organic TFTs, small molecules (pentacene or oligothiophenes) have shown by far the best electrical performance, including carrier mobilities of $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and more¹⁵⁻¹⁸. (We note that our attempts to make pentacene TFTs on monolayer dielectrics with a linear endgroup (octadecyltrichlorosilane, OTS) failed, presumably

because the interaction between the pentacene and the CH_3 -terminated SAM was so severe that the molecular ordering in the pentacene layer was insufficient to support carrier transport. That the Vuillaume group demonstrated sexithiophene TFTs with carboxyl-terminated monolayer dielectrics¹⁰ may suggest that oligothiophenes are less prone to penetration than pentacene, although they tend to provide much lower mobility.)

To understand the nature of the monolayers better, we performed a detailed scanning probe microscopy (both atomic force (AFM) and scanning tunnelling microscopy (STM)) analysis of our samples. Tapping-mode AFM was used to characterize the sample flatness and density and to image the SAM defects. We found that both methyl-terminated and phenoxy-terminated monolayers were molecularly flat over very large areas. Whereas the methyl-terminated SAM presented a small density of holes ($\sim 10 \mu\text{m}^{-2}$), no defect was observed in the phenoxy monolayers. Using STM we

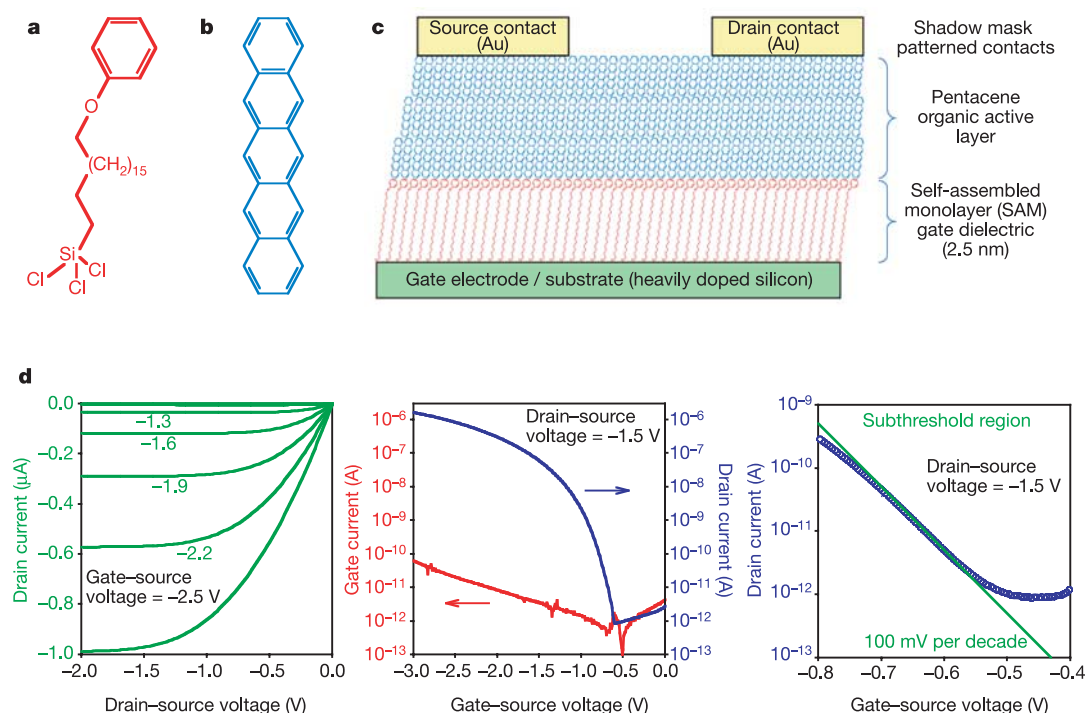


Figure 1 Chemical structures of organic materials, and cross-section and electrical characteristics of a TFT with molecular SAM gate dielectric. **a**, Structure of (18-phenoxyoctadecyl)trichlorosilane (PhO-OTS). **b**, Structure of the organic semiconductor pentacene. **c**, Cross-section of a pentacene TFT with SAM dielectric and

source/drain contacts deposited through a shadow mask. **d**, Output characteristics; **e**, transfer characteristics; **f**, subthreshold region, showing a subthreshold swing of 100 mV per decade.

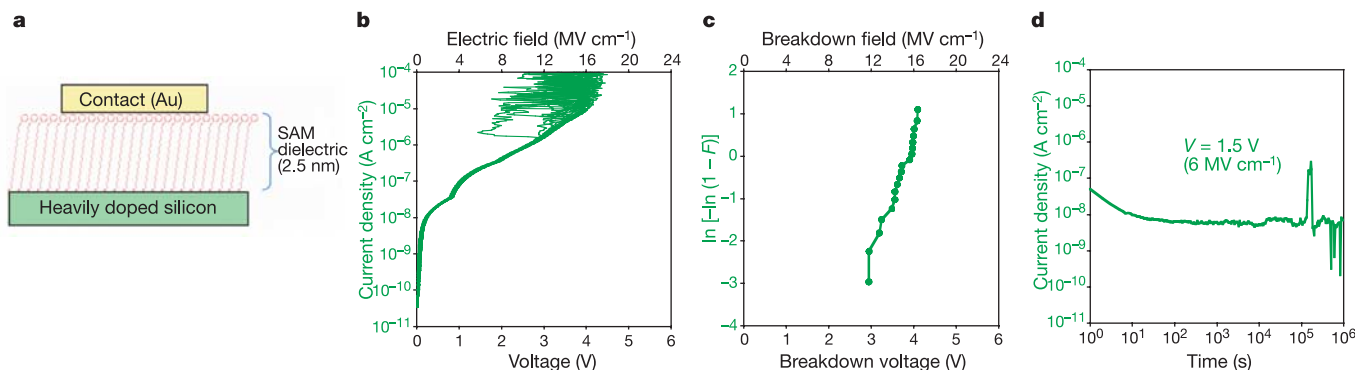


Figure 2 Electrical characteristics of Si/SAM/Au capacitors. **a**, Cross-section. **b**, Current density as a function of voltage for twenty capacitors. **c**, Breakdown field distribution. **d**, Current density as a function of time during continuous bias stressing.

found that PhO-OTS SAMs have good short-range order but no long-range order, and thus are free of domain boundaries (see Supplementary Information).

Thus, the PhO-OTS monolayer is completely pinhole-free and molecularly flat over a large area. Additionally, it is mostly amorphous in that there is no long-range order in the conjugated head groups, showing that there are no domain boundaries. This 'glassy' organic surface, composed of randomly oriented benzene rings, allows the ordering of small organic molecules. The organic nature of the interface between the insulating SAM and the molecular semiconductor may provide for a more efficient transistor channel, possibly owing to a better matching in surface energy of these materials. (We note that the best electrical performances in state-of-the-art organic TFTs were obtained on organic–organic interfaces, such as amorphous polymers, SAM-treated or polymer-treated inorganic dielectrics^{15–18}.)

To prepare TFTs with a PhO-OTS gate dielectric, a heavily doped silicon wafer was used as the substrate and gate electrode. A sufficiently large density of hydroxyl (OH) groups was created on the silicon surface by briefly exposing the substrate to an oxygen plasma. Molecular self-assembled monolayers were deposited either at room temperature from an anhydrous toluene solution, or at increased temperature from the vapour phase in a low-pressure inert ambient. Upon self-assembly, the static contact angle typically increased from 10° to 95° ± 5°, confirming the formation of a dense monolayer. The SAM dielectrics have a thickness of about 2.5 nm (slightly less than the length of the molecule), a permittivity of about 2.5 (ref. 19), and a capacitance of $9 \times 10^{-7} \text{ F cm}^{-2}$. After allowing the SAM dielectric to self-assemble on the substrate, a 30-nm-thick layer of pentacene was deposited by thermal evaporation in vacuum, and devices were completed by thermally evaporating gold through a shadow mask to define the source and drain contacts.

Figure 1 shows the schematic cross-section and the electrical characteristics of a TFT fabricated as described above. The transistor has a channel length of 130 μm, a channel width of 170 μm, a field-

effect mobility of $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (measured at a gate-source voltage of –2.5 V), a threshold voltage of –1.3 V, an on/off current ratio of 10^6 , and a subthreshold swing of 100 mV per decade. To our knowledge, this is the best subthreshold swing reported for an organic transistor, similar to those of state-of-the-art silicon MOSFETs²⁰, and close to the theoretical room-temperature limit of 60 mV per decade: $kT/[q \times \log(e)]$. The mobility of $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ is close to the best mobilities reported for organic TFTs and several orders of magnitude larger than the mobility reported by the Vuillaume group¹⁰. At a gate voltage of 1 V the gate current density is about $10^{-9} \text{ A cm}^{-2}$, much lower than the gate current density previously reported for fully processed transistors¹⁰. The drain current exceeds the gate current by a factor of 4×10^4 , confirming the excellent gate insulation provided by the top-linked monolayer dielectric. Device yield for arrays of twenty transistors is usually 100%, and the standard deviation of the device parameters is typically about ±20%. TFTs with α,α'-diethylsexithiophene (Et-6T-Et) organic semiconductor layers, identical in construction with the pentacene TFTs above, have a field-effect mobility of $0.2 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, measured at a gate-source voltage of –1.3 V (details in Supplementary Information). This value is close to the largest mobility reported for this semiconductor on an amorphous polymer dielectric¹⁷ and demonstrates the potential of using the PhO-OTS SAM dielectric for a wide range of organic semiconductors to prepare high-performance devices for low-voltage applications.

Figure 2 shows the breakdown characteristics of our SAM gate dielectric, determined on 20 Si-SAM-Au capacitors, each with a contact area of $3 \times 10^{-4} \text{ cm}^2$. These capacitors were prepared by evaporating gold contacts through a shadow mask directly onto the SAM dielectric layer. At a voltage of 1 V applied across the SAM (corresponding to an electric field of 4 MV cm^{-1}) we measure a current density of $(8 \pm 1) \times 10^{-8} \text{ A cm}^{-2}$. At 2.5 V the current density is $(8 \pm 0.5) \times 10^{-7} \text{ A cm}^{-2}$. Dielectric breakdown occurs at $14 \pm 2 \text{ MV cm}^{-1}$. These numbers for current density and breakdown voltage are comparable to or better than those of silicon

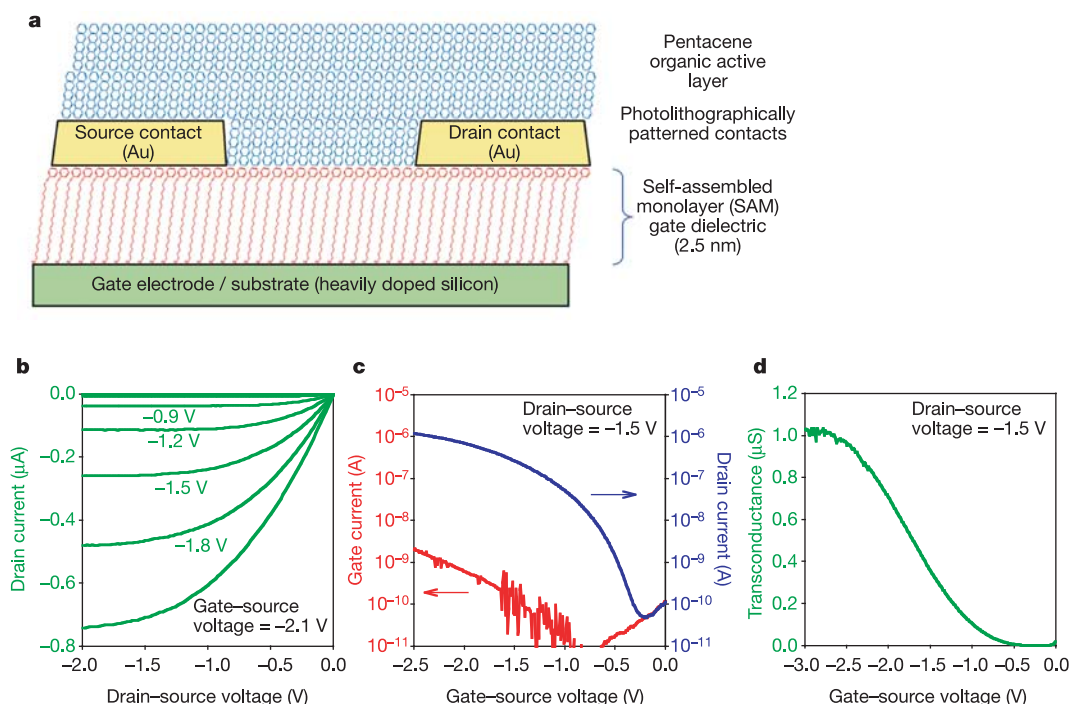


Figure 3 Cross-section and electrical characteristics of a pentacene TFT with molecular SAM dielectric and photolithographically patterned source–drain contacts. **a**, Cross-section of the TFT. **b**, Output characteristics (channel length, 5 μm; channel width,

100 μm). **c**, Transfer characteristics. **d**, Transconductance as a function of gate–source voltage, showing a maximum transconductance of 1 μS.

dioxide dielectrics of similar thickness. For example, for 3.5-nm-thick SiO₂, Sekine *et al.* reported a current density of 10⁻⁶ A cm⁻² (at 2.5 V) and a breakdown field of 14.5 MV cm⁻¹ (ref. 21). Lo *et al.* have modelled electron tunnelling through SiO₂ dielectrics and reported a current density of 10⁻⁸ A cm⁻² (at 2.5 V) for 3.5-nm-thick SiO₂ and a current density of 10⁻² A cm⁻² (at 2.5 V) for 2.5-nm-thick SiO₂ (ref. 22). These results suggest that for very thin dielectrics, molecular monolayers may provide better performance than inorganic oxides. (We note that capacitors fabricated without SAM, that is, with gold contacts evaporated directly on the plasma-activated silicon, show leakage current densities greater than 100 A cm⁻², confirming that the contribution of the native oxide film to the insulation characteristics is minimal and that the formation of a high-quality molecular SAM is essential.)

We have also carried out bias stress measurements on Si/SAM/Au capacitors. For a stress period of 10⁶ s, the data show only a slight degradation in the insulating properties of the monolayer, indicating a charge-to-breakdown of greater than 0.01 C cm⁻², similar to SiO₂ (see Fig. 2).

To integrate TFTs into circuits and displays it is often necessary to define the source and drain contacts by photolithography. The schematic cross-section of a pentacene TFT with a PhO-OTS gate dielectric and photolithographically defined contacts is shown in Fig. 3. After depositing the SAM on the activated silicon substrate, gold was thermally evaporated and patterned by photolithography and wet chemical etching. Following the gold etch, the photo resist was stripped in acetone. No special precautions were taken to protect the SAM. The contact angle changed from initially 98° after SAM deposition, to 91° after the Au etch, to 93° after stripping the photo resist, showing that the SAM was not damaged by the wet chemical processing. Transistors were completed by thermally evaporating the pentacene active layer. Figure 3 shows the electrical characteristics of a TFT with a channel length of 5 µm and a channel width of 100 µm. The transistor has a threshold voltage of -0.7 V, a subthreshold swing of 140 mV per decade, a carrier mobility of 0.05 cm² V⁻¹ s⁻¹, and a transconductance of 0.01 µS µm⁻¹, to our knowledge the largest transconductance reported for an organic transistor.

Our results show that large operating voltages are not an intrinsic feature of organic transistors, and that the operating voltage and power dissipation of organic devices can be dramatically reduced by exploiting the self-assembly of silane-based molecular dielectrics. With a thickness of 2.5 nm, these dielectrics provide a gate capacitance near 1 µF cm⁻², so that the TFTs can be operated with voltages of 2 V or less, and yet sustain gate fields of 14 MV cm⁻¹ and show gate current densities as low as 10⁻⁹ A cm⁻². In addition, the SAM dielectrics are sufficiently robust to allow the use of standard semiconductor manufacturing methods (including photolithography and metal etching), which may even permit their use as a gate dielectric in advanced silicon CMOS technology. □

Received 5 March; accepted 31 August 2004; doi:10.1038/nature02987.

1. Rogers, J. A. *et al.* Paper-like electronic displays: Large-area rubber-stamped plastic sheets of electronics and microencapsulated electrophoretic inks. *Proc. Natl Acad. Sci. USA* **98**, 4835–4840 (2001).
2. Sheraw, C. D. *et al.* Organic thin-film transistor-driven polymer-dispersed liquid crystal displays on flexible polymeric substrates. *Appl. Phys. Lett.* **80**, 1088–1090 (2002).
3. Huitema, H. E. A. *et al.* Plastic transistors in active-matrix displays. *Nature* **414**, 599 (2001).
4. Crone, B. K. *et al.* Organic oscillator and adaptive amplifier circuits for chemical vapor sensing. *J. Appl. Phys.* **91**, 10140–10146 (2001).
5. Bartic, C., Campitelli, A. & Borghs, G. Field-effect detection of chemical species with hybrid organic/inorganic transistors. *Appl. Phys. Lett.* **82**, 475–477 (2003).
6. Crone, B. *et al.* Large-scale complementary integrated circuits based on organic transistors. *Nature* **403**, 521–523 (2000).
7. de Leeuw, D. M. *et al.* Polymeric integrated circuits: fabrication and first characterisation. *2002 Int. Electron Device Meeting (IEDM) Tech. Dig.*, 293–296 (2002).
8. Baude, P. *et al.* Pentacene-based radio-frequency identification circuitry. *Appl. Phys. Lett.* **82**, 3964–3966 (2003).
9. Klauk, H. *et al.* Pentacene organic transistors and ring oscillators on glass and on flexible polymeric substrates. *Appl. Phys. Lett.* **82**, 4175–4177 (2003).

10. Collet, J., Tharaud, O., Chapoton, A. & Vuillaume, D. Low-voltage, 30 nm channel length, organic transistors with a self-assembled monolayer as gate insulating films. *Appl. Phys. Lett.* **76**, 1941–1943 (2000).
11. Boulass, C., Davidovits, J. V., Rondelez, F. & Vuillaume, D. Suppression of charge carrier tunneling through organic self-assembled monolayers. *Phys. Rev. Lett.* **76**, 4797–4800 (1996).
12. Collet, J. & Vuillaume, D. Nano-field effect transistor with an organic self-assembled monolayer as gate insulator. *Appl. Phys. Lett.* **73**, 2681–2683 (1998).
13. Schütz, M. *Self-Assembled Monolayers on Oxidized Surfaces—Anti-Corrosion and Insulation Behavior* PhD thesis, 192–193, Univ. Stuttgart (2002).
14. Lenfant, S., Krzeminski, C., Delerue, C., Allan, G. & Vuillaume, D. Molecular rectifying diodes from self-assembly on silicon. *Nano Lett.* **3**, 741–746 (2003).
15. Kelley, T. W. *et al.* High-performance OTFTs using surface-modified alumina dielectrics. *J. Phys. Chem. B* **107**, 5877–5881 (2003).
16. Klauk, H. *et al.* High-mobility polymer gate dielectric pentacene thin film transistors. *J. Appl. Phys.* **92**, 5259–5263 (2002).
17. Halik, M. *et al.* Relationship between molecular structure and electrical performance of oligothiophene organic thin film transistors. *Adv. Mater.* **15**, 917–922 (2003).
18. Kelly, T. W. *et al.* High performance organic thin film transistors. *MRS Symp. Proc.* **771**, 169–179 (2003).
19. Fontaine, P. *et al.* Octadecyltrichlorosilane monolayers as ultrathin gate insulating films in metal-insulator-semiconductor devices. *Appl. Phys. Lett.* **62**, 2256–2258 (1993).
20. Thompson, S. *et al.* A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD, and 1 µm² SRAM cell. *2002 Int. Electron Device Meeting (IEDM) Tech. Dig.*, 61–64 (2002).
21. Sekine, K., Saito, Y., Hirayama, M. & Ohmi, T. Highly reliable ultrathin silicon oxide film formation at low temperature by oxygen radical generated in high-density krypton plasma. *IEEE Trans. Electr. Dev.* **48**, 1550–1555 (2000).
22. Lo, S. H., Buchanan, D. A., Taur, Y. & Wang, W. Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFETs. *IEEE Electr. Dev. Lett.* **18**, 209–211 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements M.S. and S.M. were supported by the PhD Fellowship of the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.H. (marcus.halik@infineon.com).

Ultra-remote stereocontrol by conformational communication of information along a carbon chain

Jonathan Clayden, Andrew Lund, Lluís Vallverdú & Madeleine Helliwell

Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Many receptors¹ and allosteric proteins² function through binding of a molecule to induce a conformational change, which then influences a remote active site. In synthetic systems, comparable intramolecular information transfer can be effected by using the shape of one part of a molecule to control the stereoselectivity of reactions occurring some distance away³. However, the need for direct communication with the reaction site usually limits such remote stereocontrol to distances of not more than about five bond lengths. Cyclic structures overcome this problem by allowing the controlling centre and the reaction site^{4,5} to approach each other, but the information transfer spans only short absolute distances. Truly remote stereocontrol can, however, be achieved with rigid compounds containing amide groups: the conformation of the amides can be controlled by stereogenic centres^{6–9} and responds to that of neighbouring amide groups^{10–12} and in turn influences stereoselective reactions¹³. This strategy has allowed remote stereocontrol spanning 8 (ref. 11) or 9 (ref. 12) bonds.

Here we demonstrate stereocontrol over a reaction taking place more than 20 bond lengths from the controlling centre, corresponding to a linear distance of over 2.5 nm. This transmission of information, achieved by conformational changes relayed through the molecule, provides a chemical model of allostery and might serve as a molecular mechanism for communicating and processing information^{14–16}.

The scheme in Fig. 1 illustrates how information about the shape of the unit at A can be transmitted to the environment at B through the conformation of a series of freely rotating units *u*, *v*, *w*, ... *z*. This transmission of information depends on the reliability of three types of interaction: (1) the interaction between A and the nearby flexible unit *u*; (2) the interaction between adjacent units *u* and *v*, *v* and *w* and so on, such that each unit can respond to the spatial orientation of its neighbours; and (3) the influence of *z* on the environment at B. In chemical terms, A could be a stereogenic centre whose shape is a consequence of its absolute stereochemistry, units *u* to *z* could be freely rotating but unsymmetrical functional groups, and B could be a prochiral reactive site, the stereoselectivity of whose reactions is governed by its local environment.

Compounds containing amide groups display all three types of interaction: stereogenic centres can govern the conformation of amides^{6–9}, information about conformation can be passed between pairs of amides^{10–12}, and amide conformation can control stereoselective reactions¹³. In aromatic tertiary amides, the plane of the aromatic ring and the rigid amide are perpendicular to each other^{17–19}. Observations on ring systems bearing more than one amide group suggest that the amides interact so as to minimize the overall dipole of the molecule^{10–12}. These features, and the fact that amides are relatively rigid and unreactive, makes aromatic tertiary amides ideal building blocks to realize the information transmission system sketched in Fig. 1.

On the basis of the known conformational preference of **1** (ref. 11) and **2** (ref. 12) we designed and synthesized bisxanthenes **3** and **4** (Fig. 2). All four compounds **1–4** prefer conformations with an *anti* arrangement of the amide carbonyl groups (shown in blue), and variable temperature nuclear magnetic resonance (VT NMR) experiments indicate that the 80:20 preference of **3** for the *anti* conformation shown is increased to a >95:5 preference for the *anti*, *anti* conformation in **4**.

Xanthene **6** and bisxanthene **8**, chiral but racemic derivatives of **1**

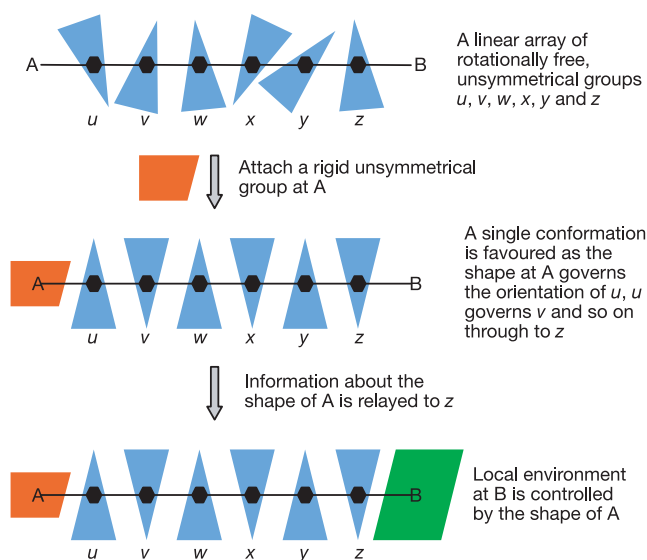


Figure 1 A strategy for conformational communication of information.

and **4**, were synthesized by the sequence of ortholithiation and palladium-catalysed coupling reactions^{20,21} shown in Fig. 3. (1*R*,2*S*)-(–)-Ephedrine condensed with the formyl groups of **6** and **8** to form the chiral oxazolidine rings²² of **9** and **10**, illustrated in red in Fig. 4. In line with earlier observations^{7,8}, the asymmetry of the oxazolidine ring should impose a right-hand (*P*) twist preference on the adjacent amide group for steric and electronic reasons. The X-ray crystal structure of **9** (see page 25 of the Supplementary Information) confirmed the preferred arrangement of the amides (blue in Fig. 4) relative to each other and relative to the oxazolidine (red in Fig. 4).

The sharp NMR spectra of **9** and **10** at 23 °C indicated that they each adopt a single conformation in solution, with the conformational preference imposed by the chiral oxazolidine ring propagated through the amide groups and thus placing the conformation of the final amide in the chain under the ultimate control of the absolute stereochemistry of the stereogenic centres in the (1*R*,2*S*)-(–)-ephedrine. The ability of this terminal amide to impose relayed stereocontrol over the remote construction of a new stereogenic centre was determined by formylating **9** and **10** and adding nucleophiles to the resultant aldehydes **11** and **12**, as illustrated in Fig. 4.

NMR (see Supplementary Figs S1 and S2) showed that, analogous to earlier work²³, addition of nucleophiles PhMgBr and EtMgBr to **11** and **12** gave the alcohols *syn*-**13** (R = Et = ethyl or R = Ph = phenyl) and *syn*-**14** (R = Et or R = Ph) as >95% of a single diastereoisomer, each of which was isolated in high yield. For comparison, authentic mixtures of *syn*- and *anti*-**13** and **14** were formed by relatively unselective addition of phenyllithium to **11** and **12** or by diluting the reaction of EtMgBr with **11** and **12** with Et₂O, which resulted in the formation of significant amounts of the *anti* diastereoisomer²³. Hydrolysis of such mixtures derived from (–)-**11** and (–)-**12** confirmed the identity of (*R*)-*anti*-**15** and (*S*)-*anti*-**16**. Thermal equilibration of (*S*)-*syn*-**15** and (*R*)-*syn*-**16** confirmed the identity of (*S*)-*anti*-**15** and (*R*)-*anti*-**16**.

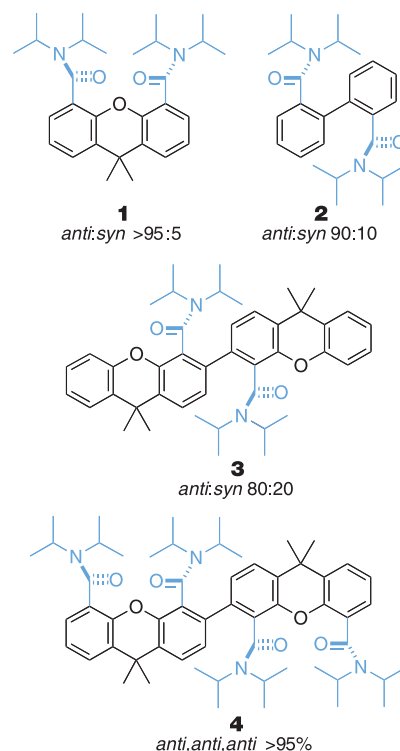


Figure 2 Conformational preference in arenedicarboxamides **1–4**.

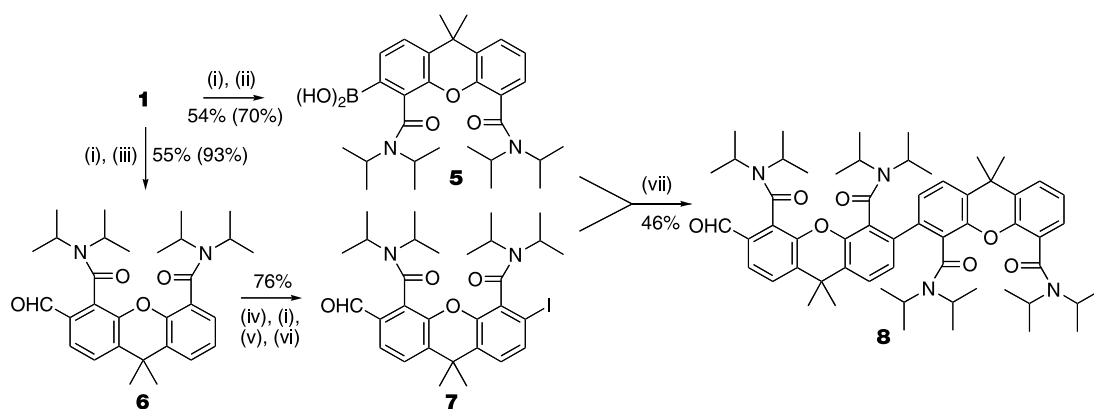


Figure 3 Synthesis of bisxanthene **8**. Reagents: (i) *n*-BuLi, N,N,N',N'-tetramethylethylenediamine (TMEDA), THF at -78°C ; (ii) $\text{B}(\text{OMe})_3$ then H_2O ; (iii) Me_2NCHO ; (iv) (1*RS*,2*SR*)-(±)-ephedrine, toluene, Δ ; (v) I_2 ; (vi) $\text{CF}_3\text{CO}_2\text{H}$, H_2O , THF;

(vii) Pd_2dba_3 (2 mol%), 9-diphenylphosphinylphenanthrene (9-DPPP) (8 mol%), $\text{Ba}(\text{OH})_2 \cdot 8\text{H}_2\text{O}$, xylene, 115°C , 50 min. (Yields in parentheses are based on recovered starting material.) Me, methyl; Δ , heat; dba, dibenzylideneacetone.

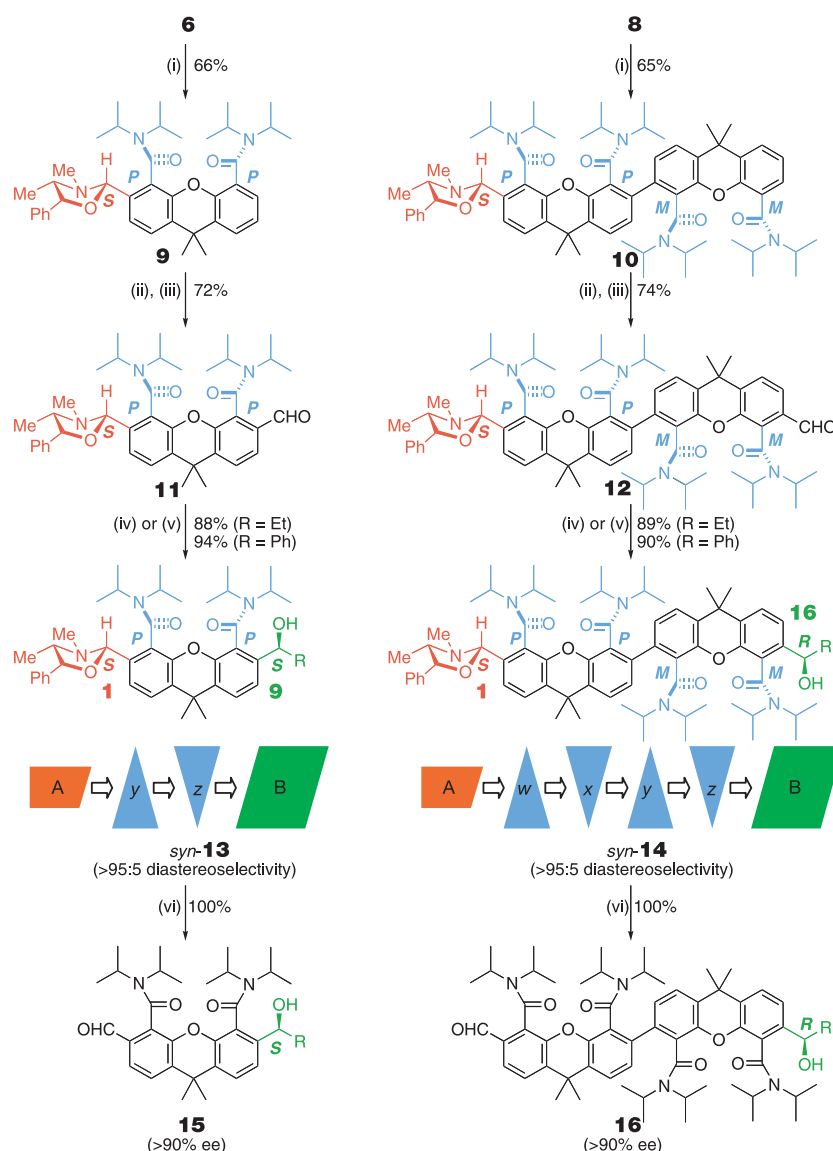


Figure 4 Remote stereocontrol in a xanthene and a bisxanthene. The influence of the oxazolidine ring shown in red (located at 'A' in the cartoon) over the orientation of the amides shown in blue (*w*, *x*, *y*, *z*) governs the formation of the new stereogenic centre

shown in green (located at 'B'). Reagents: (i) (1*R*,2*S*)-(-)-ephedrine, toluene, Δ ; (ii) *n*-BuLi (**9**) or *s*-BuLi (**10**), TMEDA, THF, -78°C ; (iii) Me_2NCHO ; (iv) EtMgBr , THF, -78°C ; (v) PhMgBr , THF, -78°C ; (vi) $\text{CF}_3\text{CO}_2\text{H}$, H_2O , THF.

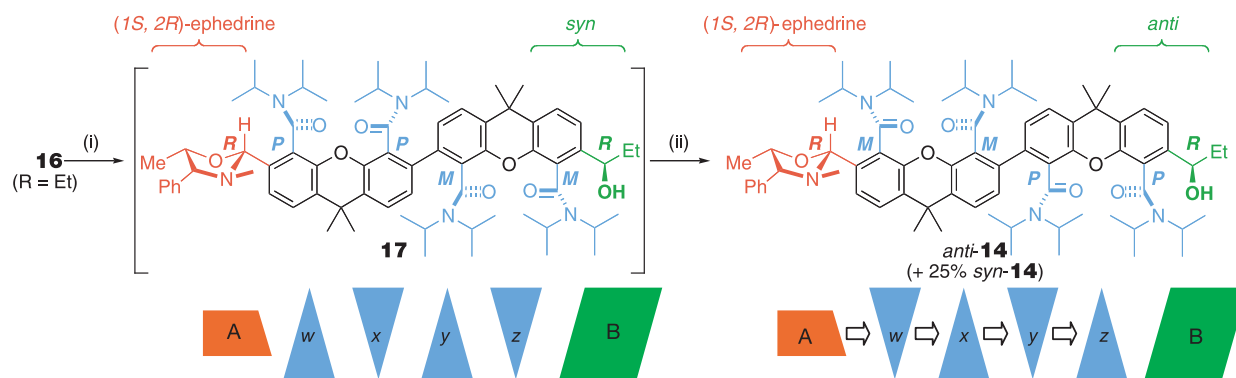


Figure 5 Inverting relative stereochemistry from a distance. Inverting the shape of A causes a mis-match with the orientation of *w*. The conformational change caused by the resulting

realignment of *w* propagates through **17** to *z*, whose reorientation influences the position of signals in the NMR spectrum at B. Reagents: (i) (1*S*,2*R*)-(+)-ephedrine; (ii) toluene, Δ .

While NMR confirmed that diastereoselective addition to the terminal formyl groups of **11** and **12** was locally diastereoselective, the efficiency of the remote transmission of information had to be confirmed by measurement of the absolute stereochemical purity of the new hydroxyl-bearing centre shown in green in Fig. 4. To this end, racemic ephedrine was used to make ($\pm$)-**9** and ($\pm$)-**10**, which were then converted to racemic ($\pm$)-*syn*-**13** and ($\pm$)-*syn*-**14**. Stirring enantiomerically pure and racemic **13** and **14** with aqueous acid cleaved the oxazolidine ring, and gentle heating partially scrambled the conformation of the amide substituents, giving enantiomerically enriched and racemic **15** and **16** as mixtures of diastereoisomers. Comparison by analytical HPLC on a chiral stationary phase (Chiralcel-ODH, 250 $\times$ 4.6 mm) of the alcohols resulting from each sequence showed that, for the alcohols **15** and **16** derived from enantiomerically pure oxazolidines, the two major diastereoisomeric peaks were accompanied by almost undetectable quantities of their enantiomers (see Supplementary Figs S4 and S5). Alcohols **15** and **16** therefore have >90% enantiomeric excess (ee): the conformation of the hydroxyl-bearing centre has been controlled by almost complete (>95:5) transmission of stereochemical information from the ephedrine-derived oxazolidine at one end of **11** and **12**, through conformationally responsive amide groups, to the other end of the molecule. Stereochemical control due to intermolecular transfer of stereochemical information (for example, complexation of one molecule of **11** or **12** to the nucleophile as it adds to another) is ruled out by the fact that nucleophilic additions to ($\pm$)-**11** and ($\pm$)-**12** are as diastereoselective as nucleophilic addition to the enantiomerically pure compounds.

The information-relaying capacity of the amides was confirmed by re-forming an oxazolidine ring from the formyl group of the diastereoisomeric mixture of aldehydes **15** and **16** using (1*S*,2*R*)-(+)-ephedrine in place of (1*R*,2*S*)-(–)-ephedrine, as illustrated for **16** (R = Et) in Fig. 5. The new, enantiomeric oxazolidine of **17** at A (shown in red) now imposes an *M* sense of twist on the first amide of the chain ('*w*' in the cartoon), which propagates through the amides to rotate the final amide ('*z*') into an *anti* relationship with the stereogenic centre at B (shown in green). The switch from *syn* to *anti* stereochemistry is clearly evident in the NMR spectra of the products (see Supplementary Figs S1 and S2). The *syn* $\rightarrow$ *anti* switches are not complete, probably because of the higher kinetic barrier to amide rotation in **13** and **14** when compared with **9** and **10**.

We find that the length over which stereochemical information can be communicated is limited by the efficiency of the synthesis of the substrates, rather than by the reliability of each stage in the information relay. The bisxanthene **10** was extended to a trisxanthene **20** by iodination and coupling with the boronic acid **19** (Fig. 6), but we were unable to synthesize a tetraxanthene by any

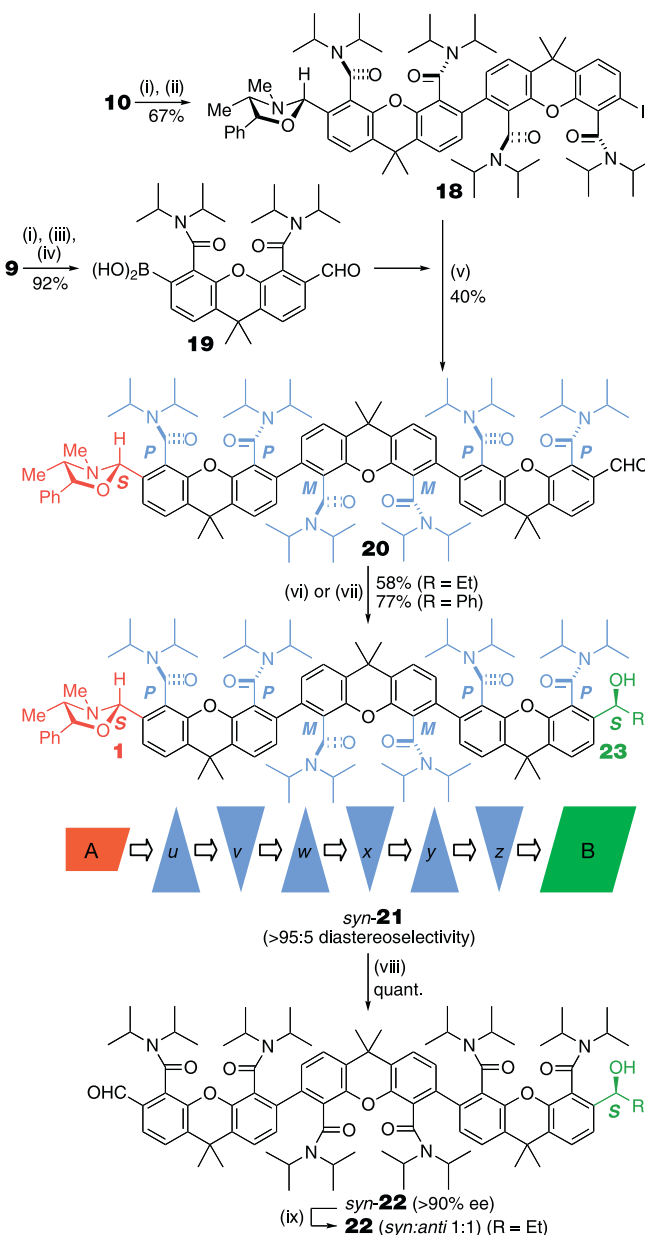


Figure 6 Ultra-remote stereocontrol in a trisxanthene. The influence of the oxazolidine ring shown in red (located at 'A' in the cartoon) over the orientation amides shown in blue (*u*, *v*, *w*, *x*, *y*, *z*) governs the formation of the new stereogenic centre shown in green (located at 'B'). Reagents: (i) *n*-BuLi, TMEDA, THF, -78°C ; (ii) I_2 ; (iii) $\text{B}(\text{OMe})_3$; (iv) $\text{CF}_3\text{CO}_2\text{H}$; (v) $\text{Pd}(\text{dba})_3$ (4 mol%), 9-DPPP (16 mol%), $\text{Ba}(\text{OH})_2 \cdot 8\text{H}_2\text{O}$, xylene, 115°C , 20 min; (vi) EtMgBr , THF, -78°C ; (vii) PhMgBr , THF, -78°C ; (viii) $\text{CF}_3\text{CO}_2\text{H}$, H_2O , THF; (ix) EtOAc , 50°C , 4 days.

reasonable coupling. The X-ray crystal structure of **20** (see page 33 of the Supplementary Information) was consistent with the expected preferred 'all *anti*' arrangement of the six amides (shown in blue in Fig. 6). Nucleophilic addition of EtMgBr and PhMgBr to **20** proceeded with >95% diastereoselectivity by NMR, giving (*S*)-*syn*-**21** (R = Et or R = Ph). Hydrolysis of the oxazolidine gave **22**, initially as >95% a single stereoisomer (by NMR and HPLC; see Supplementary Figs S3 and S6) and, after heating at 50 °C for 4 days, as a 1:1 mixture of *syn* and *anti* diastereoisomers. Determination of stereochemical purity at the new stereogenic centre of **22** (R = Ph) by HPLC and comparison with ($\pm$)-**22** (Supplementary Fig. S6) indicated (1,23)-asymmetric induction: stereochemical information is transmitted over 22 bond lengths, spanning a total distance in excess of 25 Å, with >95% reliability.

To date, the most remote stereocontrol using relayed conformational interactions has been (1,10)-asymmetric induction¹², so the (1,23)-asymmetric induction in **20** constitutes a significant advance, particularly given the modular construction of the oligo-xanthenes. Examples of (1,8-9)-, (1,10-12)-, (1,13-14)- and (1,15)-asymmetric induction (refs 24, 25, 5 and 26, and 27, respectively) are known, but these generally involve intramolecular coordination of the reaction site to the source of stereochemical information. Such coordination is impossible in **11**, **12** and **20**, where the xanthene monomers can only rotate relative to one another. The reaction site is therefore spatially remote from the source of

stereochemical information, and is not merely separated by many intervening bonds.

It is evident from the HPLC traces (Supplementary Figs S4–S6) that alcohols **15**, **16** and **22** are enantiomerically pure. Assignment of the absolute configuration at their hydroxyl-bearing centres was made by indirect methods on the basis that there is (1) a preference for an *anti* conformational relationship between the amides in related simpler compounds^{11,12} that is also evident in the X-ray crystal structures of **11** and **20**, and (2) reliable chelation-controlled *syn* selectivity in the addition of Grignard reagents to simple 2-formylnaphthamides²³. Further confirmation of the preferred stereochemistry of the aldehydes **11**, **12** and **20** is obtained by condensing them with a second equivalent of (1*R*,2*S*)-(-)-ephedrine. The oligoxanthene portion of **11** and **20** should exhibit overall C₂ symmetry, which would be 'matched' by incorporation of two homochiral ephedrine units and thus allow **23** and **25** to adopt, enantioselectively, single C₂ symmetric conformations. Indeed, both **11** and **20** condense cleanly with (1*R*,2*S*)-(-)-ephedrine to yield symmetrical products (Fig. 7). In contrast, the bisxanthene portion of **12** with its odd number of biaryl linkages is expected to prefer overall S₂ symmetry, like **2**, **3** and **4** (which all have a centre of symmetry), and therefore to be achiral. The two homochiral ephedrine moieties of **24** are mismatched with this symmetry, and a symmetrical diastereoisomer or **24** would be expected to form only with difficulty. When this condensation

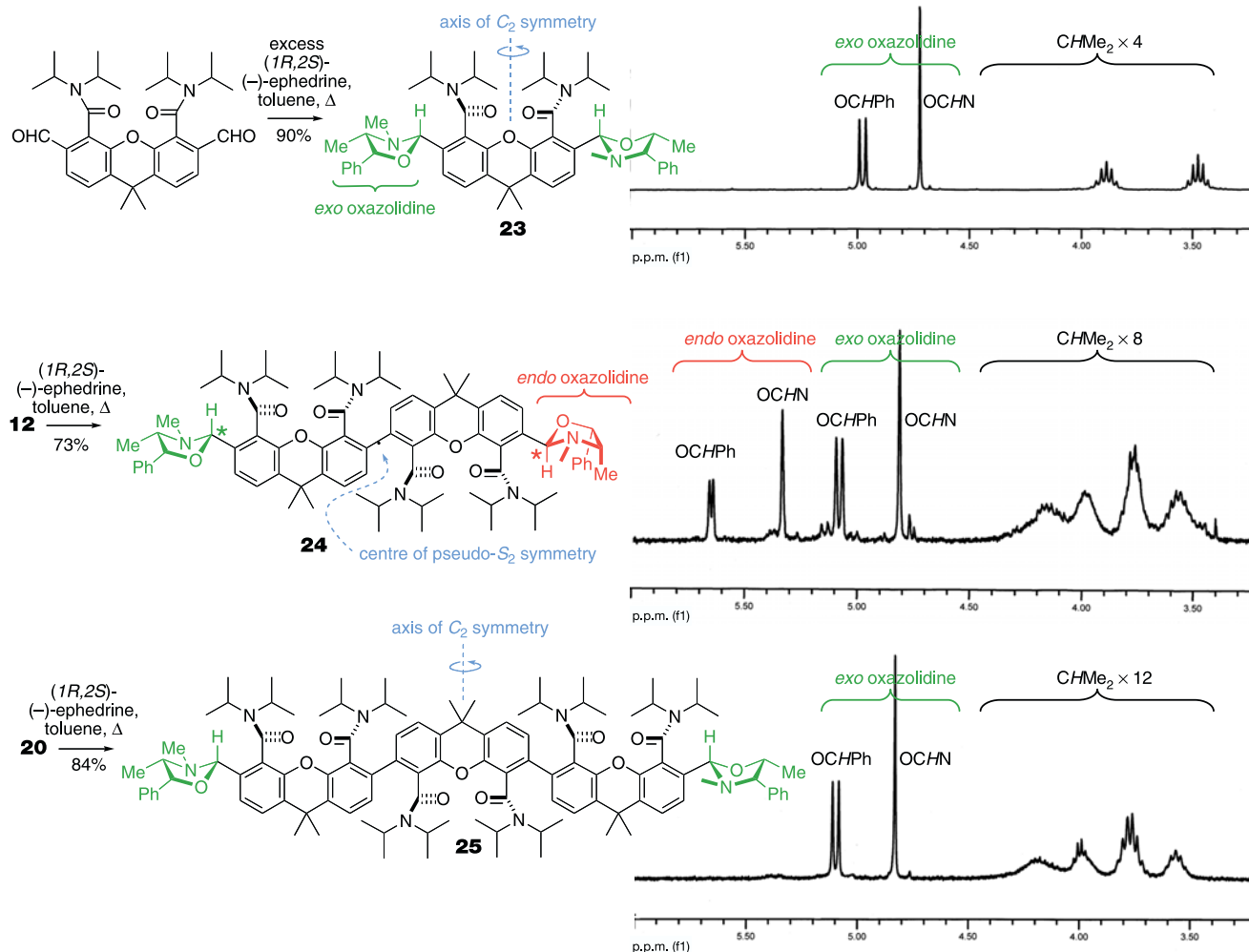


Figure 7 Symmetry matching in bis-oxazolidines. Information about the shape of one end of the molecule is relayed to the other, leading to a symmetry match in the axially chiral **23** and **25** but a symmetry mis-match in the pseudo-centrosymmetric bis-xanthene **24**.

was carried out, the bisoxazolidine **24** formed was unsymmetrical, with one of the oxazolidine rings epimeric with the other. The pseudo- S_2 symmetry of **24** forces the two oxazolidine centres (marked with asterisks in Fig. 7) to seek pseudo-enantiomeric conformations, which can only happen if one oxazolidine adopts unusual *endo* relative stereochemistry²⁸, allowing it to retain the favourable oxazolidine–amide interaction.

Extending information transmission in these systems beyond the 2.5 nm reported here will depend on the efficacy of additional coupling reactions: the conformational relay itself seems remarkably robust, with no significant degradation of information quality even after six relay steps. The synthetic versatility of amides, and in particular their ability to act as branch points, may make related compounds valuable components of future molecular devices for the transmission and processing of information^{15,16}. □

Received 26 May; accepted 10 August 2004; doi:10.1038/nature02933.

- Krauss, G. *Biochemistry of Signal Transduction and Regulation* (Wiley-VCH, Weinheim, 1999).
- Perutz, M. *Mechanisms of Cooperativity and Allosteric Regulation in Proteins* (Cambridge Univ. Press, Cambridge, 1990).
- Elie, E. L. & Wilen, S. H. *Stereochemistry of Organic Compounds* Ch. 12 (Wiley, New York, 1994).
- Mikami, K., Shimizu, M., Zhang, H.-C. & Maryanoff, B. E. Acyclic stereocontrol between remote atom centers via intramolecular and intermolecular stereo-communication. *Tetrahedron* **57**, 2917–2951 (2001).
- Linnane, P., Magnus, N. & Magnus, P. Induction of molecular asymmetry by a remote chiral group. *Nature* **385**, 799–801 (1997).
- Clayden, J., Mitjans, D. & Youssef, L. H. Lithium-sulfoxide-lithium exchange for the asymmetric synthesis of atropisomers under thermodynamic control. *J. Am. Chem. Soc.* **124**, 5266–5267 (2002).
- Clayden, J. & Lai, L. W. (–)-Ephedrine as an auxiliary for the asymmetric synthesis of atropisomeric amides by dynamic resolution under thermodynamic control. *Tetrahedron Lett.* **42**, 3163–3166 (2001).
- Clayden, J., Lai, L. W. & Helliwell, M. Using amide conformation to “project” the stereochemistry of a (–)-ephedrine-derived oxazolidine: a pair of pseudoenantiomeric chiral amido-phosphine ligands. *Tetrahedron* **12**, 695–698 (2001).
- Clayden, J. & Lai, L. W. Enantioselective synthesis of atropisomeric amides by dynamic resolution: thermodynamic control with a proline-derived diamine resolving agent. *Angew. Chem. Intl Edn* **38**, 2556–2558 (1999).
- Clayden, J., Pink, J. H. & Yasin, S. A. Conformationally interlocked amides: remote asymmetric induction by mechanical transfer of stereochemical information. *Tetrahedron Lett.* **39**, 105–108 (1998).
- Clayden, J., Kenworthy, M. N. & Youssef, L. H. Axial chirality in xanthene-4,5-dicarboxamides: 1,9-stereocontrol mediated by remote interactions between conformationally constrained amide groups. *Tetrahedron Lett.* **41**, 5171–5174 (2000).
- Clayden, J., Lund, A. & Youssef, L. H. Conformational preference and remote (1,10) stereocontrol in biphenyl-2,2′-dicarboxamides. *Org. Lett.* **3**, 4133–4136 (2001).
- Clayden, J. Stereocontrol with rotationally restricted amides. *Synlett*, 810–816 (1998).
- Krauss, R. & Koert, U. Molecular signal transduction by conformational transmission. *Synlett*, 598–608 (2003).
- Balzani, V., Credi, A. & Venturi, M. The bottom-up approach to molecular-level devices and machines. *Chem. Eur. J.* **8**, 5524–5542 (2002).
- Sauvage, J.-P. Transition-metal containing rotaxanes and catenanes in motion: toward molecular machines and motors. *Acc. Chem. Res.* **31**, 611–619 (1998).
- Cuyegkeng, M. A. & Mannschreck, A. Chromatographic separation of enantiomers and barriers to enantiomerization of axially chiral aromatic carboxamides. *Chem. Ber.* **120**, 803–809 (1987).
- Bowles, P. et al. Atroposelectivity in the reactions of ortholithiated tertiary amides with aldehydes. *J. Chem. Soc. Perkin Trans. 1*, 2607–2616 (1997).
- Ahmed, A. et al. Barriers to rotation about the chiral axis of tertiary aromatic amides. *Tetrahedron* **54**, 13277–13294 (1998).
- Snieckus, V. Directed ortho metalation. Tertiary amide and O-carbamate directors in synthetic strategies for polysubstituted aromatics. *Chem. Rev.* **90**, 879–933 (1990).
- Yin, J., Rainka, M. P., Zhang, X.-X. & Buchwald, S. L. A highly active Suzuki catalyst for the synthesis of sterically hindered biaryls: novel ligand coordination. *J. Am. Chem. Soc.* **124**, 1162–1163 (2002).
- Clayden, J., Lai, L. W. & Helliwell, M. Dynamic resolution of atropisomeric amides using proline-derived imidazolidines and ephedrine-derived oxazolidines. *Tetrahedron* **60**, 4399–4412 (2004).
- Clayden, J., McCarthy, C., Westlund, N. & Frampton, C. S. Atroposelective attack of nucleophiles on 2-formyl-1-naphthamides and their derivatives: chelation and non-chelation control. *J. Chem. Soc. Perkin Trans. 1*, 1363–1378 (2000).
- Date, M. et al. Efficient 1,8- and 1,9-asymmetric inductions in the Grignard reaction of δ - and ϵ -keto esters of 1,1′-binaphthalen-2-ols with an oligoether tether as the 2′-substituent: application to the synthesis of (–)-malyngolide. *J. Chem. Soc. Perkin Trans. 1*, 645–653 (2001).
- Tamai, Y. et al. Efficient 1,8- to 1,12-asymmetric induction in Grignard reactions of ω -keto esters by using BINOL or its 2′-oligoether derivatives as the chiral auxiliary. *J. Chem. Soc. Perkin Trans. 1*, 1141–1142 (1999).
- Magnus, N. & Magnus, P. 1,13- and 1,14-asymmetric induction; remote control. *Tetrahedron Lett.* **38**, 3491–3494 (1997).
- Heinemann, C. & Demuth, M. Short biomimetic synthesis of a steroid by photoinduced electron transfer and remote asymmetric induction. *J. Am. Chem. Soc.* **121**, 4894–4895 (1999).

28. Agami, C. & Rizk, T. Kinetic control of asymmetric induction during oxazolidine formation from (–)-ephedrine and aromatic aldehydes. *Tetrahedron* **41**, 537–540 (1985).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to the EPSRC for support of this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to J. C. (clayden@man.ac.uk). The Cambridge Crystallographic Database deposition numbers for X-ray crystal structures are, for **9**, 244658 and for **20**, 244659.

Extreme climate of the global troposphere and stratosphere in 1940–42 related to El Niño

S. Brönnimann^{1,2}, J. Luterbacher³, J. Staehelin¹, T. M. Svendby⁴, G. Hansen⁵ & T. Svenøge⁶

¹Institute for Atmospheric and Climate Science, ETH Zürich, Hönggerberg HPP, CH-8093 Zürich, Switzerland

²Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

³NCCR Climate and Institute of Geography, University of Bern, Hallerstrasse 12, CH-3012 Bern, Switzerland

⁴Department of Physics, University of Oslo, PO Box 1048, Blindern, N-0316 Oslo, Norway

⁵Norwegian Institute for Air Research, Polar Environmental Centre, Hjalmar Johansens Gate 14, N-9296 Tromsø, Norway

⁶Norwegian Polar Institute, Ny-Ålesund, PO Box 505, N-9171 Longyearbyen, Norway

Although the El Niño/Southern Oscillation phenomenon is the most prominent mode of climate variability¹ and affects weather and climate in large parts of the world, its effects on Europe and the high-latitude stratosphere are controversial^{2–5}. Using historical observations and reconstruction techniques, we analyse the anomalous state of the troposphere and stratosphere in the Northern Hemisphere from 1940 to 1942 that occurred during a strong and long-lasting El Niño event. Exceptionally low surface temperatures in Europe and the north Pacific Ocean coincided with high temperatures in Alaska. In the lower stratosphere, our reconstructions show high temperatures over northern Eurasia and the north Pacific Ocean, and a weak polar vortex. In addition, there is observational evidence for frequent stratospheric warmings and high column ozone at Arctic and mid-latitude sites. We compare our historical data for the period 1940–42 with more recent data and a 650-year climate model simulation. We conclude that the observed anomalies constitute a recurring extreme state of the global troposphere–stratosphere system in northern winter that is related to strong El Niño events.

Scientists in the early 1940s observed unusually high values of total ozone over several sites in Europe, but did not present an explanation. At the same time, exceptional climatic conditions were registered at the Earth's surface, but were never analysed in a large-scale context. Moreover, a prolonged El Niño occurred in 1939–42, raising the question of a possible relation between El Niño, European climate, and the northern stratosphere⁶. In order to study this period in detail, we have compiled, digitized and re-evaluated all available total ozone observations^{7–9} and several tens of thousands of geopotential height (GPH) and temperature profiles from radio-

sondes and aircraft from 1939–44¹⁰. The latter served to statistically reconstruct monthly mean fields of GPH and temperature for the northern extratropics up to the 100 hPa level in the lower stratosphere (around 16 km altitude)¹¹. Here we analyse these data and compare the results to more recent data and a climate model simulation.

The El Niño event started in autumn 1939, reached full strength in January 1940 and lasted, with varying intensity, until spring 1942 (Supplementary Fig. 1). The characteristic anomalies in the northern extratropics described below began in December 1939 or January 1940 and ended between February and April 1942, with a considerable persistence during this period (Supplementary Fig. 1). In the following we focus on the extended winter period (January to April; the season most indicative of large-scale coupling processes) of the years 1940, 1941 and 1942. Averaged anomaly fields are shown in Fig. 1 (top). In order to assess the magnitude of the individual anomaly features, corresponding time series for the twentieth century are shown in Fig. 2, together with the El Niño index NINO3.4, the Pacific Decadal Oscillation index (PDOI) and the North Atlantic Oscillation index (NAOI; for definitions, see Methods).

The dominant global feature was the contrast between high tropical and low extratropical sea surface temperatures (SSTs) in both hemispheres, most pronounced in the Pacific (reflected by exceptional peaks in NINO3.4 and PDOI). Surface air temperatures (SATs) were exceptionally high in Alaska, Canada and central Asia (Fig. 1), but low in Siberia and extremely low in central and northeastern Europe, where the three severe winters in a row (T_{Euro} in Fig. 2) played an important role in the Second World War. A similar, large-scale temperature pattern also appeared in summer and autumn 1940 and 1941, though weaker in magnitude (Supplementary Fig. 2). The dominant features in the wintertime sea level pressure field (SLP, Fig. 1) were a weak Icelandic low and an intensified Aleutian low; their anomaly difference (IL–AL in Fig. 2) was twice as large in 1940 and 1941 than at any other time in the

twentieth century. The weak Icelandic low and Azores high led to a strongly negative NAOI. Positive SLP deviations were observed over Scandinavia and North America, and negative deviations over Labrador and Manchuria. Distinct anomaly patterns are also found in the reconstructed upper-level fields (see Methods). The 300 hPa GPH field reflects an anomaly in planetary wave structure; positive deviations appear over Canada and Greenland, negative deviations over the western North Pacific, parts of the North Atlantic, and central Europe. At the Earth's surface and in the troposphere, the period 1940–42 represents an extreme climatic anomaly of hemispheric to global extent.

Re-evaluated upper-air data and reconstructions¹¹ reveal that this is also the case in the stratosphere. The wintertime 100 hPa GPH anomaly field (Fig. 1 top) shows a large difference between the polar region and the mid-latitudes (Z100 in Fig. 2), indicating a weak and meridionally expanded polar vortex. Temperatures at 100 hPa exhibit a positive anomaly over northern Eurasia and the North Pacific, and negative anomalies over the northern Atlantic and Canada. The data suggest a frequent occurrence of major midwinter stratospheric warmings (MMW, at least in January 1940, February 1941 and February 1942)¹². The anomaly in stratospheric circulation was accompanied by large total ozone deviations. A strong positive anomaly around 1940–42 appears in all six available ozone series (Fig. 3), at sites spread throughout the Northern Hemisphere. The same is found for lower-quality series from Oxford, UK¹³, and Table Mountain, California¹⁴. The highest annual mean values in the series from Arosa in Switzerland (since 1926) and Tromsø in the Norwegian Arctic (since 1935) both occurred in 1940 (Fig. 2; note that anthropogenic ozone depletion affects total ozone after around 1970). Chemical effects can largely be excluded—rather, the total ozone data point to an extreme perturbation of ozone transport, and hence circulation, in the stratosphere.

Analysed in context of twentieth-century climate variability, the 1940–42 period stands out as an anomaly unique in strength, but weaker anomalies of the same type are also found after 1948. There

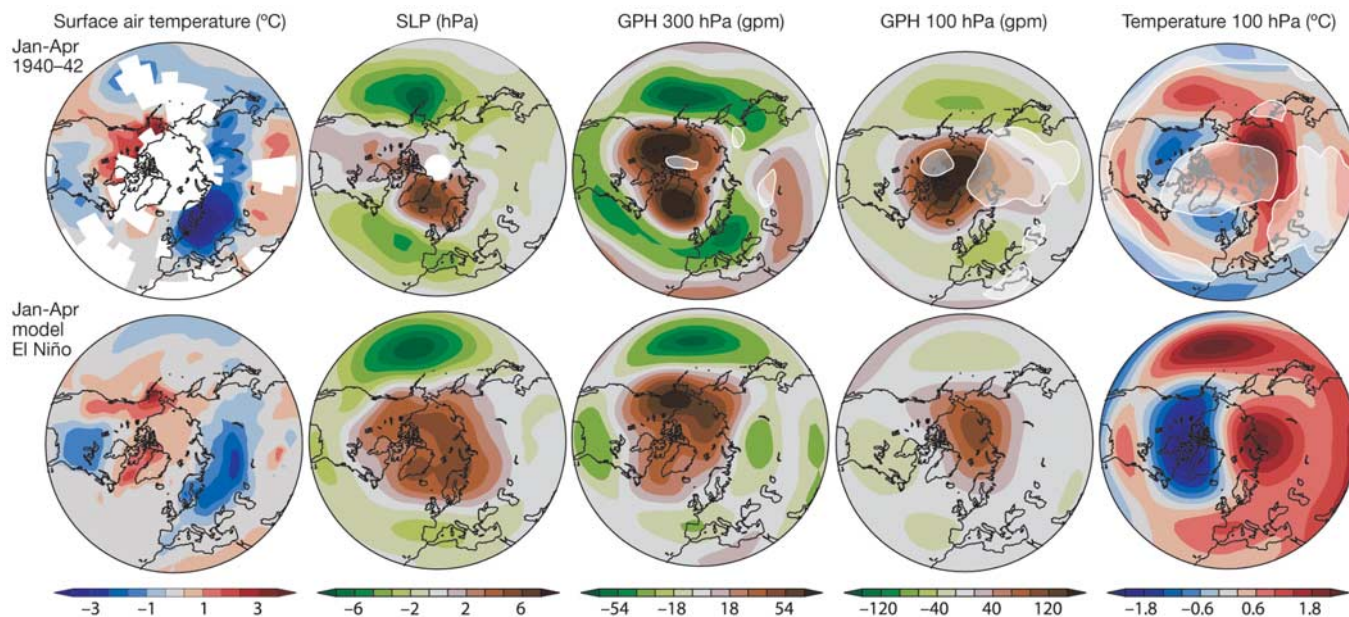


Figure 1 Consistent northern winter climate extremes related to strong El Niño events. Top, polar stereographic projection of anomaly fields (with respect to 1961–90) of surface air temperature²⁷, sea level pressure (SLP)²⁹, 300 hPa and 100 hPa geopotential height (GPH in units of geopotential metres, gpm) and 100 hPa temperature¹¹, in the northern extratropics averaged from January to April, 1940–42.

Light shadings denote a low reconstruction skill (reduction of error RE < 0.2 based on two split sample validations¹¹). Bottom, same anomaly fields for the winters during the 26 strongest El Niño years (October–September NINO3.4 > 1 °C) in the CCSM-2.0 control run.

are seven cases where all index series relating to the Atlantic-European sector and the stratosphere (NAOI, T_{Euro} , Z100, total ozone at Arosa and Tromsø), after removing an 11-yr moving average to account for trends and possible solar cycle effects, have the same sign as in 1940–42. All of these cases—the most pronounced were 1969/70, 1987 and 1998—concur with El Niño conditions (positive NINO3.4), which therefore are considered as a possible driving factor.

The anomalies in the Pacific region can at least partly be attributed to El Niño, which is known to affect the North Pacific and surrounding areas via changes in the Hadley circulation and Rossby wave generation^{15,16}. Possible low-frequency oceanic processes and ocean–atmosphere feedbacks in the North Pacific might also have contributed to the cold central North Pacific and strong Aleutian low in 1940/41 and 1987^{17,18}. As to the Atlantic-European sector, the anomalies at the surface and in the stratosphere are

dynamically consistent with each other. The wintertime relation between the Icelandic low and the polar vortex and their effect on European climate and total ozone are well documented^{19–21}. To what extent European winter climate and the northern stratosphere are affected by El Niño is a matter of debate^{2–5}. According to several observational and modelling studies, the ‘canonical’ El Niño winter signal in Europe consists of cold temperatures in Northern Europe, high SLP from Iceland to Scandinavia and low SLP over central and western Europe^{2,4,22}, as in 1940–42. Also, strong El Niños have been found to be associated with a weak polar vortex and more frequent major stratospheric warmings^{5,6}. However, not all El Niños show these features and other studies find no consistent signal, which may be due to the strong variability of extratropical circulation, the small number of strong events, disturbing effects of volcanic eruptions and anthropogenic influences, interference with the Quasi Biennial Oscillation in the stratosphere^{5,6}, or a varying, non-stationary signal^{4,16}. In summary, all anomalies found in 1940–42 are qualitatively consistent with El Niño influence.

To further explore the concurrence of extremes in Pacific and European climate and the northern stratosphere and their possible relation to El Niño, we used a 650-yr control run of a coupled climate model (CCSM-2.0) and calculated the same index series as in Fig. 2. Extremes in the different series tend to be concomitant. Selecting only strong El Niño years (October–September average of NINO3.4 > 1 °C, 26 cases in 650 yr), significant deviations of the same sign as in 1940–42 appear in all series (Fig. 2 right, the selected year is denoted 0). The mean wintertime anomaly fields for these cases (Fig. 1 bottom) show a strong similarity to those of the early 1940s, not only with respect to SAT and SLP, but also the planetary wave structure and strength of the polar vortex, as evident in 300 and 100 hPa GPH. Even for 100 hPa temperature we find the same characteristic pattern. In addition, the selected El Niño years exhibit very similar SAT patterns as in 1940 and 1941 for summer and, to a lesser extent, autumn (Supplementary Fig. 2). These results confirm that, first, the concurrence of the characteristic anomalies in 1940–42 represents a consistent extreme state of the global troposphere–stratosphere system on interannual timescales. Second, they suggest that this extreme state can be related to strong El Niños.

The CCSM-2.0 model does not simulate stratospheric ozone. Nevertheless, analysing further dynamical properties can give important indications with respect to stratosphere–troposphere coupling and stratospheric ozone. We find a significant increase

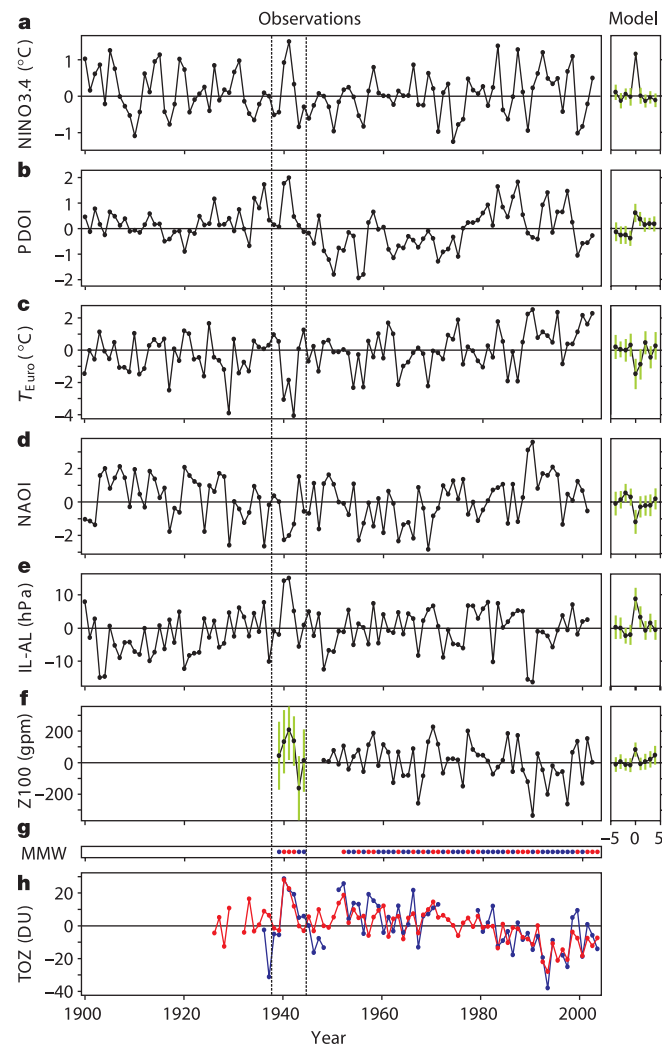


Figure 2 Climate indices and total ozone series from 1900 to 2003 (left) and composite series for the 26 strongest El Niño years (preceding and following 4 yr are also shown) in the CCSM-2.0 control run (right). **a**, NINO3.4 (ref. 26); **b**, PDOI (ref. 28); **c**, T_{Euro} ; **d**, NAOI; **e**, IL-AL; **f**, Z100; **g**, occurrence (red) or absence (blue) of major midwinter (December–February) stratospheric warmings (MMW)^{12,31}; and **h**, total ozone at Arosa⁷ (red) and Tromsø (blue). **a**, **b** and **h** are annual averages, **c–f** are January–April averages (see Methods). Green bars give 95% confidence intervals for reconstructions (left, **f**) and mean values (right, **a–f**). Dashed lines denote the 1938–44 period.

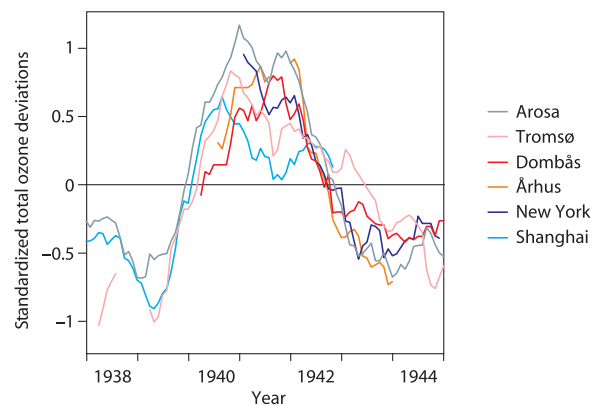


Figure 3 The 1940–42 total ozone anomaly. Standardized monthly anomalies (with respect to the 1938–44 mean annual cycle) of total ozone at Arosa (Switzerland)⁷, Tromsø, Dombås⁸ (both Norway), Århus (Denmark)⁹, New York (USA)⁹ and Shanghai (China)⁹ smoothed with a 12-month moving average.

in the zonally averaged, early-to-midwinter (November–February) meridional eddy heat flux at 100 hPa (45°–75° N) in the selected El Niño years, suggesting that more upward propagating planetary wave activity reached the stratosphere. A stronger wave activity flux diminishes the strength of the polar vortex and strengthens the meridional circulation in the middle stratosphere^{23,24}. It leads to more ozone transport from the tropical source regions to the extratropics²⁴, stronger descent over the polar area and hence a warm lower stratosphere²³ and high Arctic total ozone in late winter²⁴. Furthermore, pulses of increased upward propagating wave activity act as trigger for major stratospheric warmings²⁵. Hence, this mechanism is qualitatively consistent with all stratospheric features of the 1940–42 anomaly (for which no heat flux can be calculated). It not only provides an explanation for the total ozone increase in the Arctic but would also contribute, together with ozone redistribution in the lower stratosphere owing to the change in planetary wave structure¹², to the total ozone increase at mid-latitudes.

The results suggest that the global climate anomaly in 1940 to 1942—previously poorly documented—constitutes a key period for our understanding of large-scale climate variability and global El Niño effects. □

Methods

Climate model

CCSM-2.0 is a coupled, global climate model with four components (ocean, atmosphere, land, sea ice). We used the control run b20.007 provided by the University Corporation for Atmospheric Research, Boulder, Colorado (for documentation see <http://www.cesm.ucar.edu/models/ccsm2.0/>). The model was run fully coupled for 370 yr in a spin-up mode with constant 1990 conditions. The final control run was continued from year 350 and was run fully coupled until year 999. The atmospheric component CAM1 was run in a T42 horizontal spectral resolution with 26 levels. Ozone concentrations were fixed to the 1990 annual cycle. Note that this includes ozone-hole conditions over Antarctica, not representative of 1940–42 conditions. However, we expect no strong effect on the circulation of the northern extratropics. We used monthly averages for the years 350–999.

Climate indices

Several climate indices were calculated based on monthly mean values. NINO3.4 (ref. 26), T_{NPAO} , T_{ALASKA} and T_{EURO} are averaged temperature anomalies for the areas 5° S–5° N/120°–170° W, 30°–50° N/155°–225° E, 50°–75° N/180°–240° E and 45°–70° N/20°–55° E, respectively (HadCRUT2v data, consisting of SAT over land areas merged with SST²⁷). The PDOI is the amplitude, defined for all calendar months, of the first principal component of monthly North Pacific SSTs (poleward of 20° N) obtained from October to March²⁸. Z100 is the difference in 100 hPa GPH anomalies between 75°–90° N and 40°–55° N, IL and AL are SLP²⁹ anomalies for the areas 60°–70° N/10°–40° W and 40°–55° N/140°–170° W, respectively. The NAOI is the difference of standardized SLP²⁹ anomalies at 64.13° N/21.9° W and 37.73° N/23.7° W. In order to obtain annual series, T_{EURO} , Z100, NAOI and IL–AL were averaged from January to April, when large-scale coupling processes are expected to be most influential^{20,23,24}. Annual averages were used for PDOI, total ozone (calendar year) and NINO3.4 (October of previous year to September, thus leading all extratropical indices by three months), because anomalies in total ozone and SST have a long persistence. Monthly meridional (transient plus stationary) eddy heat flux was calculated from the model data as $\langle v'T \rangle + \langle v' \rangle \langle T' \rangle$, where v and T denote instantaneous meridional wind and temperature, respectively, $\langle \rangle$ denotes a monthly average, and primes and stars denote deviations from monthly and zonal averages, respectively. It was subsequently zonally averaged between 45° and 75° N from November to February. The chosen lead of two months with respect to Z100 maximizes the correlation with Z100 ($r = 0.75$), consistent with other studies²³. Upper-level fields were reconstructed from historical radiosonde and aircraft data as well as SAT station data and SLP by means of principal component regression models calibrated in the 1948–94 period¹¹. They were supplemented with NCEP/NCAR reanalysis³⁰ after 1948. Unless otherwise noted, anomalies refer to the 1961–90 mean annual cycle (years 350–999 in the CCSM-2.0 data). For the determination of spatial or temporal averages, 50% of the data must be available. Spatial averages from gridded data were obtained using the grid cell areas as weights. Tests and 95% confidence intervals for mean values were calculated assuming a t -distribution.

Total ozone

The Tromsø total ozone data 1935–49 were homogenized partially by introducing empirical correction factors. The measurements before 15 August 1939 performed with an older instrument (Fery spectrograph) were homogenized with the later measurements (Dobson spectrophotometer) based on simultaneous measurements in the period August 1939 to September 1940 (scaling factor of 1.37). Zenith sky measurements 1940–49 were adjusted by calculating ratios of pure direct-sun/zenith-blue daily means and daily means including zenith-cloud measurements, yielding consistent correction factors (T.S. and G.H., manuscript in preparation). After November 1978, the record was supplemented

with TOMS Version 7 overpass data in order to fill gaps. For all series, monthly means were calculated if at least six daily values were available in order to include winter months with sparse data (most monthly means are based on more than 15 values).

Received 16 April; accepted 27 August 2004; doi:10.1038/nature02982.

1. Diaz, H. & Markgraf, V. (eds) *El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts* (Cambridge Univ. Press, Cambridge, UK, 2000).
2. Fraedrich, K. & Müller, K. Climate anomalies in Europe associated with ENSO extremes. *Int. J. Climatol.* **12**, 25–31 (1992).
3. Pozo-Vázquez, D., Esteban-Parra, M. J., Rodrigo, F. S. & Castro-Diez, Y. The association between ENSO and winter atmospheric circulation and temperature in the North Atlantic region. *J. Clim.* **14**, 3408–3420 (2001).
4. Graetbatch, R. J., Lu, J. & Peterson, K. A. Nonstationary impact of ENSO on Euro-Atlantic winter climate. *Geophys. Res. Lett.* **31**, doi:10.1029/2003GL018542 (2004).
5. van Loon, H. & Labitzke, K. The Southern Oscillation. Part V: The anomalies in the lower stratosphere of the Northern Hemisphere in winter and a comparison with the Quasi-Biennial Oscillation. *Mon. Weath. Rev.* **115**, 357–369 (1987).
6. Labitzke, K. G. & van Loon, H. *The Stratosphere. Phenomena, History, and Relevance* (Springer, Berlin, 1999).
7. Stachelin, J. *et al.* Total ozone series of Arosa (Switzerland). Homogenization and data comparison. *J. Geophys. Res.* **103**, 5827–5841 (1998).
8. Svendby, T. M. Reanalysis of total ozone measurements at Dombås and Oslo, Norway, from 1940 to 1949. *J. Geophys. Res.* **108**, doi:10.1029/2003JD003963 (2003).
9. Brönnimann, S., Cain, J. C., Stachelin, J. & Farmer, S. F. G. Total ozone observations prior to the IGY. II: Data and quality. *Q. J. R. Meteorol. Soc.* **129**, 2819–2843 (2003).
10. Brönnimann, S. A historical upper-air data set for the 1939–1944 period. *Int. J. Climatol.* **23**, 769–791 (2003).
11. Brönnimann, S. & Luterbacher, J. Reconstructing Northern Hemisphere upper-level fields during World War II. *Clim. Dyn.* **22**, 499–510 (2004).
12. Brönnimann, S., Luterbacher, J., Stachelin, J. & Svendby, T. M. An extreme anomaly in stratospheric ozone over Europe in 1940–1942. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL019611 (2004).
13. Goldsmith, P., Tuck, A. F., Foot, J. S., Simmons, E. L. & Newson, R. L. Nitrogen oxides, nuclear weapon testing, Concorde and stratospheric ozone. *Nature* **244**, 545–551 (1973).
14. Angione, R. J. & Roosen, R. G. Baseline ozone results from 1923 to 1955. *J. Clim. Appl. Meteorol.* **22**, 1377–1383 (1983).
15. Alexander, M. A. *et al.* The atmospheric bridge: Influence of ENSO teleconnections on air-sea interaction over the global oceans. *J. Clim.* **15**, 2205–2231 (2002).
16. Trenberth, K. E. *et al.* Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.* **103**, 14291–14324 (1998).
17. Chavez, F. P., Ryan, J., Lluch-Cota, S. E. & Niquen, C. M. From anchovies to sardines and back: Multidecadal change in the Pacific Ocean. *Science* **299**, 217–221 (2003).
18. Trenberth, K. E. & Hurrell, J. W. Decadal atmosphere-ocean variations in the Pacific. *Clim. Dyn.* **9**, 303–319 (1994).
19. Ambaum, M. H. P. & Hoskins, B. J. The NAO troposphere-stratosphere connection. *J. Clim.* **15**, 1969–1978 (2002).
20. Hurrell, J., Kushnir, Y., Ottens, G. & Visbeck, M. (eds) *The North Atlantic Oscillation. Climatic Significance and Environmental Impact* (AGU, Washington DC, 2003).
21. Stachelin, J., Harris, N. R. P., Appenzeller, C. & Eberhard, J. Ozone trends: A review. *Rev. Geophys.* **39**, 231–290 (2001).
22. Merkel, U. & Latif, M. A high resolution AGCM study of the El Niño impact on the North Atlantic/European sector. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL013726 (2002).
23. Newman, P. A., Nash, E. R. & Rosenfield, J. What controls the temperature of the arctic stratosphere during the spring? *J. Geophys. Res.* **106**, 19999–20010 (2001).
24. Randel, W. J., Wu, F. & Stolarski, R. Changes in column ozone correlated with the stratospheric EP flux. *J. Meteorol. Soc. Jpn* **80**, 849–862 (2002).
25. Limpasuvan, V., Thompson, D. W. J. & Hartmann, D. L. The life cycle of the Northern Hemisphere stratospheric sudden warmings. *J. Clim.* **17**, 2584–2597 (2004).
26. Kaplan, A. *et al.* Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.* **103**, 18567–18589 (1998).
27. Jones, P. D. & Moberg, A. Hemispheric and large-scale surface air temperature variations: An extensive revision and update to 2001. *J. Clim.* **16**, 206–223 (2003).
28. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific decadal climate oscillation with impacts on salmon. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
29. Trenberth, K. E. & Paolino, D. A. The Northern Hemisphere sea level pressure data set: Trends, errors, and discontinuities. *Mon. Weath. Rev.* **108**, 855–872 (1980).
30. Kistler, R. *et al.* The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.* **82**, 247–268 (2001).
31. Labitzke, K. G. & Naujokat, B. The lower Arctic stratosphere in winter since 1952. *SPARC Newsl.* **15**, 11–14 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements S.B. was funded by the Swiss National Science Foundation, the Holderbank Foundation and the Janggen-Poehn Foundation. J.L. was funded by the Swiss National Science Foundation (NCCR climate). Surface temperature and NCEP/NCAR reanalysis data were provided by Climatic Research Unit, (Norwich, UK) and the NOAA-CIRES Climate Diagnostics Center (Boulder, USA), respectively. SLP and CCSM-2.0 data were provided by NCAR/UCAR (Boulder, USA). We thank P. Della-Marta for English corrections.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S. B. (stefan.broennimann@env.ethz.ch).

Release of gold-bearing fluids in convergent margin magmas prompted by magnetite crystallization

Weidong Sun^{1,2}, Richard J. Arculus³, Vadim S. Kamenetsky^{1,4} & Raymond A. Binns^{3,5}

¹Max-Planck-Institut für Chemie, Mainz, Postfach 3060, Mainz 55020, Germany

²Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, Australia

³Department of Earth and Marine Sciences, Australian National University, Canberra, ACT 0200, Australia

⁴Centre for Ore Deposit Research and School of Earth Sciences, University of Tasmania, GPO Box 252-79, Hobart, Tasmania 7001, Australia

⁵CSIRO Exploration and Mining, North Ryde, NSW 1670, Australia

A relationship between convergent margin magmas and copper–gold ore mineralization has long been recognized^{1–6}. The nature of the genetic link is controversial, particularly whether the link is due to high-oxygen-fugacity (f_{O_2}) melts and fluids released from subducted slabs^{5–7} or to brine exsolution during magmatic evolution⁴. For submarine, subduction-related volcanic glasses from the eastern Manus basin, Papua New Guinea, we here report abrupt decreases in gold and copper abundances, coupled with a switch in the behaviour of titanium and iron from concentration increases to decreases as SiO_2 rises. We propose that the abrupt depletion in gold and copper results from concurrent sulphur reduction as a result of f_{O_2} buffering, causing enhanced formation of copper–gold hydrosulphide complexes that become scavenged from crystallizing melts into cogenetic magmatic aqueous fluids. This process is particularly efficient in oxidized arc magmas with substantial sulphate. We infer that subsequent migration and cooling of exsolved aqueous fluids create links between copper–gold mineralization and arc magmatism in the Manus basin^{8,9}, and at convergent margins in general^{1–6}.

Most of the world's copper–gold (Cu–Au) ore deposits (for example, epithermal/porphyry types) are associated with convergent margin magmas^{1–7}, usually characterized by a high f_{O_2} (ref. 6). Although these ore metals might ultimately have been recycled from the subducted slab⁶, the subarc mantle is not significantly enriched in Au and Cu (ref. 7), and so far the relationship between Cu–Au mineralization and subduction-derived magmas remains controversial. Many different genetic models have been developed to explain this relationship^{4–7}, building mainly on the fact that subduction zone magmas have a relatively high f_{O_2} . These models include the possibilities that slab-released oxidizing fluids are capable of concentrating metals in the sulphide-bearing metasomatic assemblages of the mantle wedge that become preferentially melted during formation of arc magma⁷, and that high f_{O_2} might promote the oxidation of residual sulphide in the wedge, liberating chalcophile elements^{3,6}. Other models favour sulphide undersaturation in the magmas⁶, thus increasing magmatic enrichment in incompatible Au and Cu during crystallization.

The behaviour of Au in evolving magma systems is not well understood¹⁰. Early studies suggested that Au is compatible during magma formation¹¹ but this is not always true¹². Gold concentrations in arc volcanic rocks, although variable, increase during the early stage of magmatic evolution (incompatible behaviour), then decrease in felsic rocks (apparently compatible)^{10,13}. One possible explanation is that under special chemical and physical conditions, a considerable amount of Au can be dissolved in H_2O -rich vapour¹⁴ and fluids^{4,7,15} and therefore potentially become lost through degas-

sing of H_2O -saturated arc magmas. The widely variable Au concentrations reported for subaerially erupted arc volcanic rocks¹⁰ might partly reflect different degrees of Au loss during degassing. This process is less effective for eruptions on the deep sea floor. In this study we analysed submarine, subduction-related volcanic glasses and olivine-hosted glass inclusions (entrapped parental melts) from the eastern Manus basin (dredged in the depth range 1,640–2,000 m) using a laser-ablation, inductively coupled plasma-source mass spectrometry (LA-ICP-MS) technique refined for quantifying trace elements at concentrations less than 1 part per 10^9 (p.p.b.)^{16–18}. These data provide an outstanding opportunity to understand the behaviour of Au and Cu during the fractionation of arc-type magmas.

Essentially aphyric eastern Manus basin glasses range in composition from basalt to rhyodacite, with SiO_2 contents from 50.4 to 74.1 wt%, and MgO contents from 11.7% to 0.3% (calculated anhydrous). They have strong arc geochemical affinities, with relative depletion in Nb and Ta and enrichment in H_2O , Cs, Rb, Ba, U and Pb (refs 18,19). Our sample set (Table 1) from a large CSIRO collection of well-characterized samples was selected as a coherent fractionation suite, evidently derived from closely comparable basaltic parents in adjacent volcanic edifices (Supplementary Figs 1 and 2). Notably, TiO_2 and FeO (total) change from

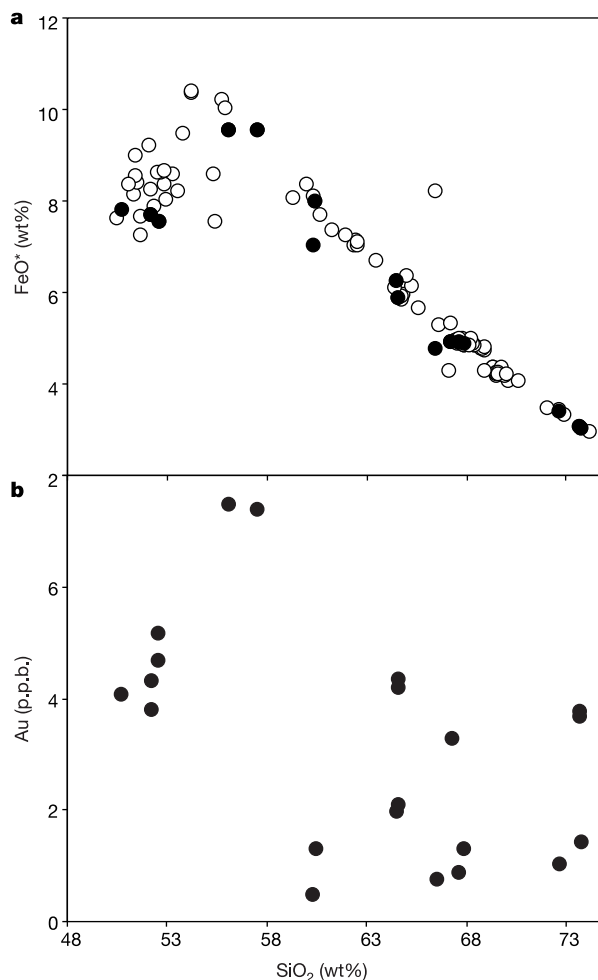


Figure 1 Selected compositional characteristics of eastern Manus basin volcanic glasses. **a**, SiO_2 wt% versus total Fe as FeO (FeO*); **b**, SiO_2 wt% versus Au. SiO_2 and FeO total are recalculated to 100% volatile-free with Fe^{3+}/Fe_{total} set at 0.25. Gold was determined with the same method as that used previously for analysing p.p.b. to sub-p.p.b. levels of Re (refs 16–18; see Supplementary Information for detail).

positive to negative correlations with SiO₂ at ~58% SiO₂ (Table 1, Fig. 1a, Supplementary Fig. 2), corresponding to a significant change in magma evolution arising from cessation of olivine crystallization and commencement of titanomagnetite crystallization. Gold concentrations in these highly vesicular Manus glasses mostly range from 0.9 to 7.8 p.p.b. (Table 1), comparable to previous data¹³ but with better sensitivity and precision. Two glass pieces include points extremely enriched in Au (up to 73 p.p.b.; Table 1), similar to previous observations on Re distribution¹⁸. However, their ablation spectrum patterns show short-lived high-Au signals that are not directly coupled with Re (Supplementary Fig. 3), suggesting that these elements are in part hosted by different trace minerals (such as sulphides) possibly within vesicles or fluid inclusions. Gold in olivine-hosted melt inclusions is high and varied (3.2–8.7 p.p.b., Table 1), spanning its abundance in the host glass (4.1 p.p.b.). Cu distribution within specific glasses is more homogeneous.

The abundance of Cu measured by LA-ICP-MS in the eastern Manus basin samples initially increases with increasing SiO₂, reaching peak values at ~58% SiO₂. Significantly, Cu concentrations drop abruptly beyond 58% SiO₂ (Table 1, Supplementary Fig. 2). Conventional bulk Cu analyses of 96 aphyric glasses from the Manus basin exhibit an identical trend. Gold concentrations, excluding anomalies mentioned above, show a similar behaviour with a distinct but less marked decrease beyond 58% SiO₂ (Fig. 1b). Total S in the glasses displays a less coherent but similar pattern, with a more gradual decrease beyond peak values of ~350 p.p.m. at ~58–64% SiO₂. In contrast, mafic melt inclusions have considerably higher and varied total S up to 1,000 p.p.m. (Supplementary Fig. 4). These indicate that even the mafic glasses have lost S, probably in the form of sulphate^{19,20}, which did not remove much, if any, Au from the magma. The Yb/Au and Yb/Cu ratios of the more mafic glasses and their melt inclusions are fairly constant (Fig. 2),

indicating that Au and Yb behave in a similarly incompatible manner during early fractional crystallization. The Yb/Au ratios of melt inclusions hosted in olivine (Fo_{89–91}) are comparable to those of the host basalt glass, precluding significant Au loss during initial fractional crystallization. The average Yb/Au (350) and Cu/Au (25,000) of mafic samples are close to but slightly lower than primitive mantle values (440 and 30,000 respectively)²¹.

By contrast, the Yb/Au and Yb/Cu values of more felsic glasses (>58% SiO₂) are variable and considerably higher than those of the mafic glasses and melt inclusions, reflecting the abrupt decreases in Au and Cu concentrations (Table 1, Fig. 2) requiring a process of removal of Au and Cu at ~58% SiO₂ during magma evolution. Low Cu and Au in felsic magmas has previously been attributed to partitioning of these metals into magnetite^{10,13}, to pre-eruptive degassing¹³ or to precipitation of magmatic sulphides⁵. Sulphide precipitation is an unlikely explanation for the eastern Manus suite given low total S contents in glasses (Supplementary Fig. 4) and given the absence or extreme rarity of sulphide globules in glasses or occasional cumulate xenoliths. Preliminary LA-ICP-MS analysis of titanomagnetite in an east Manus basin andesite (<6 p.p.b. Au, 2.9 p.p.m. Cu; see also ref. 10) indicates that extraction of this mineral will not account for the marked diminution of these metals in glasses at ~58% SiO₂. All of the magmas (mafic to felsic) from the eastern Manus basin are volatile-saturated, as shown by abundant fluid inclusions in olivine and fluid bubbles in glasses^{19,22}. Consequently, sustained pre-eruptive degassing²³ does not on its own produce the marked change in behaviour of Au and Cu at ~58% SiO₂, coinciding with the onset of magnetite crystallization.

Gold and Cu are both highly chalcophile elements whose behaviour is controlled by the abundance and oxidation state of S, for example, as a HS-bearing complex in fluids²⁴. They can also partition into Cl-rich aqueous fluids, as a Cl-bearing complex²⁵. Given the high abundance of Cl in arc magmas, a Cl complex might

Table 1 Selected major and trace element compositions of Manus glasses and melt inclusions

Sample no.	<i>n</i> †	SiO ₂ † (wt%)	TiO ₂ (wt%)	FeO* (wt%)	Cu† (p.p.m.)	1σ	Yb† (p.p.m.)	1σ	S (p.p.m.)	Au (p.p.b.)	1σ
MD6	3	73.68	0.36	3.06	17.4	0.4	4.73	0.09	0	1.47	0.11
MD39	3	72.55	0.39	3.46	19.4	0.2	4.31	0.03	67	1.06	0.17
87DR-I	1	73.53	0.43	3.12	16.0	0.2	4.25	0.06	70	3.70	0.20
87DR-II	1	73.53	0.43	3.12	17.7	1.0	4.46	0.02		73.5	18.2
87DR-III	1	73.53	0.43	3.12	18.3	0.7	4.40	0.02		3.81	0.19
MD28	3	67.76	0.67	4.91	19.6	0.1	3.89	0.09	37	1.35	0.06
MD65	3	67.53	0.67	4.95	18.9	0.1	3.39	0.05	23	0.92	0.15
MD114	3	67.14	0.75	4.97	28.1	0.3	3.58	0.01	77	3.31	0.19
MD28B†	5	66.38	0.82	4.80	20.1	1.2	3.6	0.1		0.79	0.56
MD36	3	64.40	0.86	6.29	25.6	2.3	4.0	0.7	85	2.02	0.61
MD53A-I	1	64.49	0.83	5.94	30.7	0.2	3.32	0.06	93	4.22	0.54
MD53A-II	1	64.49	0.83	5.94	32.2	0.1	3.47	0.03		4.37	0.54
MD53A-III	1	64.49	0.83	5.94	28.2	0.2	3.52	0.02		2.13	0.18
MD36B†	5	60.19	0.97	7.09	35.2	7.9	3.2	0.3		0.51	0.05
MD53B	3	60.35	1.01	8.05	34.6	0.9	3.4	0.1	190	1.33	0.16
MD7	2	57.42	1.05	9.60	234	1.3	2.73	0.02	150	7.42	0.6
86DR	3	56.01	0.75	9.58	197	2.8	2.38	0.05	93	7.51	0.24
MD101A	3	52.47	0.40	7.59	108	0.5	1.41	0.05	20	4.73	0.75
MD101B	3	52.47	0.40	7.59	117	1.6	1.67	0.06	67	5.20	0.79
MD3-I	1	52.08	0.41	7.75	100.7	0.5	1.29	0.01	47	3.84	0.21
MD3-II	1	52.08	0.41	7.75	102.6	0.3	1.37	0.01		4.34	0.21
MD3-III	1	52.08	0.41	7.75	104.1	0.6	1.39	0.02		31.3	9.9
MD38†	5	50.60	0.42	7.87	108	3	1.12	0.01	108	4.10	0.20
Melt inclusions in olivine (MD-38)§											
M5D					191	4	1.74	0.016		6.2	0.2
M5D'					223	22	1.68	0.018		4.7	0.3
M6D					267	42	2.45	0.028		7.2	0.6
M6D'					173	9	2.45	0.02		8.7	0.9
M7C					180	3	1.86	0.023		4.3	0.2
M7C'					159	9	1.67	0.019		3.9	0.2
M11C					114	3	1.06	0.014		3.2	0.2
M12B					112	2	1.04	0.014		3.3	0.2

Major elements were analysed by X-ray fluorescence recalculated to 100% volatile-free with Fe³⁺/Fe_{total} set at 0.25. The major elements of melt inclusions have not been analysed; an estimated 12 wt% CaO was adopted for LA-ICP-MS data reduction¹⁹, which might introduce offsets but has no effects on ratios¹⁷. *n*, number of analyses. Trace elements are averages when *n* ≥ 2. FeO*, total Fe as FeO. †Data from ref. 18.

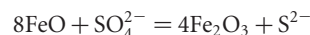
‡Major elements from ref. 19.

§Heterogeneous samples; each row represents one analysis on an individual piece of glass.

be important in scavenging Au and Cu out of the magmas but there are no experiments so far to demonstrate which complex is more important at magmatic temperature and pressures. Experiments at 100 MPa, in a high-chloride (total Cl = 2 molal) and low-H₂S system show that with increasing temperature AuCl₂⁻ becomes the dominant species²⁵. In contrast, experiments at 50 MPa show that with increasing temperature the solubility of Au increases as a HS complex^{24,26}. In crustal magma chambers at low pressures, a HS complex is therefore likely to be the main species during early stages of magma evolution. This hypothesis is supported by other observations. The high CO₂ concentrations^{19,20,22} and *f*_{O₂} characteristic of Au deposits in the eastern Manus basin magmas indicate that Au⁺ is the favoured ionic state²⁷. As one of the softest cations, Au⁺ forms stronger complexes with soft ligands (such as HS⁻) than with moderately hard anions such as Cl⁻ (ref. 27). Consequently, we propose Au and Cu abundances in the Manus basin melts and associated fluids to be largely governed by reduced S complexes.

The cogenetic Manus basin melts from basalt to rhyodacite have a relatively constant Fe³⁺/Fe total of ~23 atom% (Supplementary Fig. 5). This implies an *f*_{O₂} uniformly about 2log₁₀ orders more oxidized than the synthetic fayalite-magnetite-quartz buffer²⁸, which is typical for igneous rocks associated with Au ore deposits⁶. A change in magmatic Ti and Fe compositional trends in the Manus basin suite (Fig. 1a, Supplementary Fig. 2) controlled primarily by the commencement of titanomagnetite (Fe²⁺Fe₂³⁺O₄-Fe₂²⁺TiO₄ solid solution) crystallization should be accompanied by changes in relative Fe redox ratio in the absence of *f*_{O₂} buffering. We propose that the constant relative redox state over the compositional range of the Manus basin magmas is paralleled by sustained redox buffering and exchange between exsolved fluid and magma. Because of the high *f*_{O₂}, sulphate dominates in fluids associated with the mafic magmas, as confirmed by the presence of sulphate minerals in fluid bubbles within olivine-hosted melt inclusions^{20,22,23}. However, the onset of magnetite fractional crystallization and increased rate of total Fe and particularly Fe³⁺ removal from the magma results in

a shift in this redox exchange²⁹:



Reduction of SO₄²⁻ to S²⁻ in the melt promotes the formation of Au and Cu hydrosulphide complexes, which efficiently partition into a coexisting high-temperature aqueous supercritical fluid^{24,26}, explaining the observed sharp depletion of Au and Cu in the felsic Manus glasses.

The escape of deep-seated Cu-Au rich magmatic fluids, derived thereby at the changeover from olivine to magnetite crystallization in the source chamber, provides the critical link between subduction-related magmas and Cu-Au ore deposits in modern and fossil arcs. Cooling at depth of upwardly mobilized fluids will create typical convergent margin porphyry/epithermal Cu-Au deposits. At upper crustal levels, ascending magmatic fluids may mix with seawater and with metals released during the hydrothermal alteration of a volcanic pile to form Cu-Au-rich massive sulphide deposits such as the seabed chimneys of the eastern Manus basin^{9,30}. □

Received 16 May; accepted 25 August 2004; doi:10.1038/nature02972.

- Richards, J. P. Petrology and geochemistry of alkaline intrusives at the Porgera gold deposit, Papua New Guinea. *J. Geochem. Explor.* **35**, 141–199 (1990).
- Mueller, D. & Groves, D. I. Direct and indirect associations between potassic igneous rocks, shoshonites and gold-copper deposits. *Ore Geol. Rev.* **8**, 383–406 (1993).
- Sillitoe, R. H. Characteristics and controls of the largest porphyry Cu-Au and epithermal Au deposits in the circum-Pacific region. *Aust. J. Earth Sci.* **44**, 373–388 (1997).
- Ulrich, T., Guether, D. & Heinrich, C. A. Gold concentrations of magmatic brines and the metal budget of porphyry copper deposits. *Nature* **399**, 676–679 (1999).
- Mueller, D., Franz, L., Herzig, P. M. & Hunt, S. Potassic igneous rocks from the vicinity of epithermal gold mineralization, Lihir Island, Papua New Guinea. *Lithos* **57**, 163–186 (2001).
- Mungall, J. E. Roasting the mantle: Slab melting and the genesis of major Au and Au-rich Cu deposits. *Geology* **30**, 915–918 (2002).
- McInnes, B. I. A., McBride, J. S., Evans, N. J., Lambert, D. D. & Andrew, A. S. Osmium isotope constraints on ore metal recycling in subduction zones. *Science* **286**, 512–516 (1999).
- Binns, R. A. & Scott, S. D. Actively forming polymetallic sulfide deposits associated with felsic volcanic rocks in the eastern Manus back-arc basin, Papua New Guinea. *Econ. Geol.* **88**, 2226–2236 (1993).
- Shipboard Scientific Party. Leg 193 Summary. In *Proceedings of the Ocean Drilling Program, Initial Reports* (eds Binns, R. A., Barriga, F. J. A. S. & Miller, D. J.) 1–84 (Ocean Drilling Program, 2002).
- Togashi, S. & Terashima, S. The behavior of gold in unaltered arc tholeiitic rocks from Izu-Oshima, Fuji, and Osoreyama volcanic areas, Japan. *Geochim. Cosmochim. Acta* **61**, 543–554 (1997).
- Keays, R. R. & Scott, R. B. Precious metals in ocean-ridge basalts; implications for basalts as source rocks for gold mineralization. *Econ. Geol.* **71**, 705–720 (1976).
- Zentilli, M., Brooks, R. R., Helgason, J., Ryan, D. E. & Zhang, H. The distribution of Au in volcanic rocks of eastern Iceland. *Chem. Geol.* **48**, 17–28 (1985).
- Moss, R., Scott, S. D. & Binns, R. A. Gold content of eastern Manus Basin volcanic rocks; implications for enrichment in associated hydrothermal precipitates. *Econ. Geol.* **96**, 91–107 (2001).
- Williams-Jones, A. E., Migdisov, A. A., Archibald, S. M. & Xiao, Z. Vapor-transport of ore metals. In *Water-Rock Interaction, Ore Deposits, and Environmental Geochemistry: A Tribute to David A. Crerar* (eds Hellman, R. & Wood, S. A.) 279–305 (Spec. Publ. Vol. 7, Geochem. Soc., St Louis, 2002).
- Loucks, R. R. & Mavrogenes, J. A. Gold solubility in supercritical hydrothermal brines measured in synthetic fluid inclusions. *Science* **284**, 2159–2163 (1999).
- Sun, W., Bennett, V. C., Eggins, S. M., Arculus, R. J. & Perfit, M. R. Rhenium systematics in submarine MORB and back-arc basin glasses: Laser ablation ICP-MS results. *Chem. Geol.* **196**, 259–281 (2003).
- Sun, W. D., Bennett, V. C., Eggins, S. M., Kamenetsky, V. S. & Arculus, R. J. Enhanced mantle-to-crust rhenium transfer in undegassed arc magmas. *Nature* **422**, 294–297 (2003).
- Sun, W. D., Arculus, R. J., Bennett, V. C., Eggins, S. M. & Binns, R. A. Evidence for rhenium enrichment in the mantle wedge from submarine arc volcanic glasses (Papua New Guinea). *Geology* **31**, 845–848 (2003).
- Kamenetsky, V. S. *et al.* Parental basaltic melts and fluids in eastern Manus backarc basin; implications for hydrothermal mineralisation. *Earth Planet. Sci. Lett.* **184**, 685–702 (2001).
- Kamenetsky, V. S. *et al.* Fluid bubbles in melt inclusions and pillow-rim glasses: high-temperature precursors to hydrothermal fluids? *Chem. Geol.* **183**, 349–364 (2002).
- McDonough, W. F. & Sun, S. S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Yang, K. & Scott, D. S. Possible contribution of a metal-rich magmatic fluid to a sea-floor hydrothermal system. *Nature* **383**, 420–423 (1996).
- Yang, K. H. & Scott, S. D. Magmatic degassing of volatiles and ore metals into a hydrothermal system on the modern sea floor of the eastern Manus basin, western Pacific. *Econ. Geol.* **97**, 1079–1100 (2002).
- Gibert, F., Pascal, M. L. & Pichavant, M. Gold solubility and speciation in hydrothermal solutions: Experimental study of the stability of hydrosulphide complex of gold (AuHS⁰) at 350 to 450°C and 500 bars. *Geochim. Cosmochim. Acta* **62**, 2931–2947 (1998).
- Gammons, C. H. & Williams-Jones, A. E. Chemical mobility of gold in the porphyry-epithermal environment. *Econ. Geol.* **92**, 45–59 (1997).
- Archibald, S. M., Migdisov, A. A. & Williams-Jones, A. E. The stability of Au-chloride complexes in water vapor at elevated temperatures and pressures. *Geochim. Cosmochim. Acta* **65**, 4413–4423 (2001).
- Phillips, G. N. & Evans, K. A. Role of CO₂ in the formation of gold deposits. *Nature* **429**, 860–863 (2004).
- Kilinc, A., Carmichael, I. S. E., Rivers, M. L. & Sack, R. O. The ferric-ferrous ratio of natural silicate liquids equilibrated in air. *Contrib. Mineral. Petrol.* **83**, 136–140 (1983).

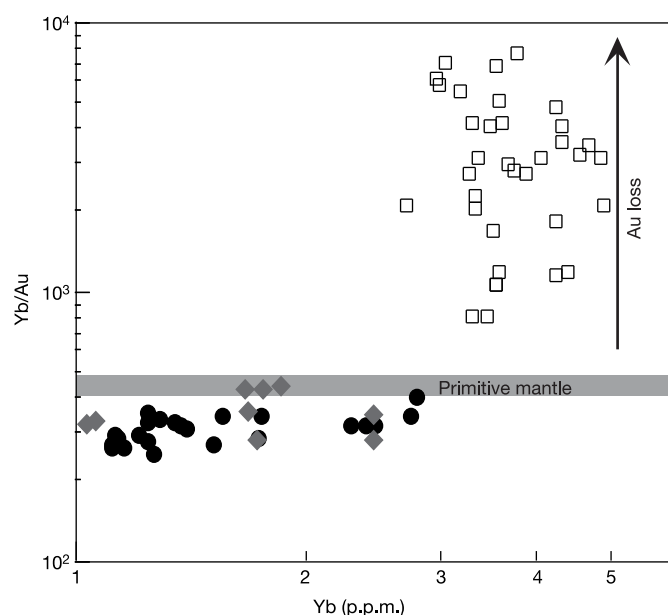


Figure 2 Plot of Yb/Au against Yb showing the relative incompatibility of Au with Yb. The fairly constant Yb/Au for mafic samples (circles) and melt inclusions (diamonds) suggests that Au is moderately incompatible, similarly to Yb. The high and varied Yb/Au of felsic samples (squares) indicates significant Au loss.

29. Carmichael, I. S. E. & Ghiorso, M. S. Oxidation–reduction relations in basic magma: a case for homogeneous equilibria. *Earth Planet. Sci. Lett.* **78**, 200–210 (1986).
30. Roberts, S. *et al.* Contrasting evolution of hydrothermal fluids in the PACMANUS system, Manus Basin: the Sr and S isotope evidence. *Geology* **31**, 805–808 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. M. Eggins, C. Allen and J. M. G. Shelley for technical support with the laser-ablation ICP-MS analyses; S.-S. Sun, A. W. Hofmann, J. E. Snow and S. J. Galer for discussions; and A. E. Williams-Jones for a review. This work was supported by a Alexander von Humboldt Research Fellowship to W.D.S. and by a Friedrich Wilhelm Bessel Award to V.S.K. R.J.A. was supported by the Australian Research Council, and R.A.B. by CSIRO and a mineral company consortium.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.J.A. (richard.arculus@anu.edu.au).

A Silurian sea spider

Derek J. Siveter^{1,2}, Mark D. Sutton¹, Derek E. G. Briggs³ & David J. Siveter⁴

¹Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

²Geological Collections, University Museum of Natural History, Oxford OX1 3PW, UK

³Department of Geology & Geophysics, Yale University, PO Box 208109, New Haven, Connecticut 06520-8109, USA

⁴Department of Geology, University of Leicester, Leicester LE1 7RH, UK

Pycnogonids (sea spiders) are marine arthropods numbering some 1,160 extant species. They are globally distributed in depths of up to 6,000 metres, and locally abundant^{1,2}; however, their typically delicate form and non-biomineralized cuticle has resulted in an extremely sparse fossil record that is not accepted universally³. There are two opposing views of their phylogenetic position: either within Chelicerata as sister group to the euechelicerates^{4–7}, or as a sister taxon to all other euarthropods⁸. The Silurian Herefordshire Konservat-Lagerstätte⁹ in England (~425 million years (Myr) BP) yields exceptionally preserved three-dimensional fossils that provide unrivalled insights into the palaeobiology of a variety of invertebrates^{10–14}. The fossils are preserved as calcitic void in-fills in carbonate concretions within a volcanoclastic horizon¹⁵, and are reconstructed digitally¹². Here we describe a new pycnogonid from this deposit, which is the oldest adult sea spider by ~35 Myr and the most completely known fossil species. The large chelate first appendage is consistent with a chelicerate affinity for the pycnogonids. Cladistic analyses place the new species near the base of the pycnogonid crown group, implying that the latter had arisen by the Silurian period.

Phylum Arthropoda
Subphylum Chelicerata
Class Pycnogonida
Haliestes dasos gen. et sp. nov.

Etymology. Greek: *Halios* (of the sea) plus *chrestes* (soothsayer; alluding to the forward-facing eye tubercle); *dasys* (hairy) plus *gloutos* (rump).

Holotype. Oxford University Museum of Natural History (OUM) C.29571. Figure 1a–j, reconstructed in three dimensions. The specimen, despite its relatively small size, is interpreted to be an adult on the basis of its completely developed appendages⁵, and its long, distally setiferous ovigers probably indicate that it is male¹. No other material is known.

Horizon and locality. Wenlock Series, Silurian period; Herefordshire, England.

Diagnosis. Pycnogonid with well-developed chelate chelicera, palp and oviger, the last two of which are subchelate; eye tubercle on anterior face of cephalosoma; walking legs subchelate and laterally flattened in the last few segments; rings or annular markings lacking on lateral processes, walking legs and trunk end (referred to as ‘abdomen’ in traditional pycnogonid terminology).

Description. The combined length of body and trunk end⁵ (anterior of cephalosoma to posterior of trunk end) is 3.5 mm. The cephalosoma is sub-pentagonal and slightly swollen where the chelicera, palp and oviger attach, and bears a low eye tubercle on its anterior face (eyes are not preserved). The lateral process is short and narrow. The chelicera (Fig. 1a–e, h) is large and robust and about 0.9 times the body length when extended. The scape consists of two segments, and is similar in length to the chela. The third segment comprises a stout palm and tapering fixed finger, and is opposed by the fourth segment, a movable finger. The fingers overlap slightly in the closed position. (This four-segmented structure contrasts with the three-segmented chelicera of most extant pycnogonids⁶).

The palp (Fig. 1a–e, g) is of similar length to the body. It includes at least eight segments (Fig. 1g). The first and second are the longest, but some segment boundaries are difficult to discern; those identified as segments 2 and 3 may be divided into two. The terminal segments form a subchela. The oviger (Fig. 1a–f), which originates ventro-laterally, is similar in form and length to the palp and comprises nine segments, the first being the longest. Straight, mostly inwardly directed setae occur distally on palp and oviger.

The proboscis (Fig. 1a–e, i) descends very steeply and slightly forwards, and is slightly shorter than the body length. It is sub-circular in cross-section, narrowest about mid-length, and terminates in a small, medial mouth. The pharyngeal lumen shows the characteristic Y-shaped transverse section¹⁶, reflecting the three longitudinal antimeres (Fig. 1c; see also Supplementary Movie). Such a section has not previously been documented in a fossil.

The first walking leg (Fig. 1a–d), comprising nine segments, is at least twice as long as the body. The leg is sub-circular in cross-section proximally, becoming more flattened in the last few segments, and terminating in a subchela. Interpretation of proximal segmentation is based partially on functional considerations (see Supplementary Note 1). The second segment is the longest whereas the seventh is shortest. The third to seventh segments bear pairs of straight, inwardly directed setae, and the eighth a single inward seta. The sixth to eighth segments also bear pairs of outwardly directed setae. The other walking legs (Fig. 1a–d) are similar to the first, but are slightly longer where completely preserved. The first segment of the fourth leg, at least, bears a pair of dorsal setae distally, and probably also proximally. Cement glands and gonopores have not been identified.

The trunk between the cephalosoma and trunk end (Fig. 1a–d) is about 0.7 times the body length. Two small, axial, forwardly directed setae or spines occur anterior to the third walking leg; similar processes are indicated weakly elsewhere on the trunk. The second to fourth lateral processes are better differentiated than the first and are directed increasingly more posteriorly, particularly the fourth.

The trunk end (Fig. 1a, b, d, j) is about 0.18 times the body length. In lateral view it comprises three ‘elements’ (segment boundaries have not been discerned); the proximal and distal elements are shorter and inclined slightly upwards and downwards respectively, whereas the central element is near horizontal. The central section bears an axial node posteriorly, with two pairs of setae: one pair directed postero-dorsally, the other postero-ventrally (Fig. 1j). Between this node and the trunk there is a suggestion of one or possibly two short, axial setae or spine bases. The position of the anus is unknown.

Discussion. Only four other fossil pycnogonid species are known: *Palaeoisopus problematicus*, *Palaeopantopus maucheri* and *Palaeothea devonica* from the Lower Devonian of Germany (~390 Myr BP)^{17–18}, and the larval *Cambropycnogon klausmuelleri* from the Upper Cambrian of Sweden (~500 Myr BP)⁴. The last three of these species are known from just five, one and seven specimens respectively. Undescribed pycnogonid material has been reported from the Jurassic period of Germany and France^{4,19}.

Haliestes has already acquired the distinctive character suite that

typifies the pycnogonids: a prominent external proboscis, modification of the third appendage into an oviger, a reduced trunk end and uniramous limbs (the fossil does not reveal the nature or position of the genital openings). This distinctive morphology underlies the difficulties in the phylogenetic placement of the group. However, the well-developed chelate chelicera in this earliest example for pycnogonids supports the inclusion of the group within the Chelicerata, as the most robust autapomorphy for the taxon^{4–7}. No truly chelate (as opposed to subchelate) appendage has yet been

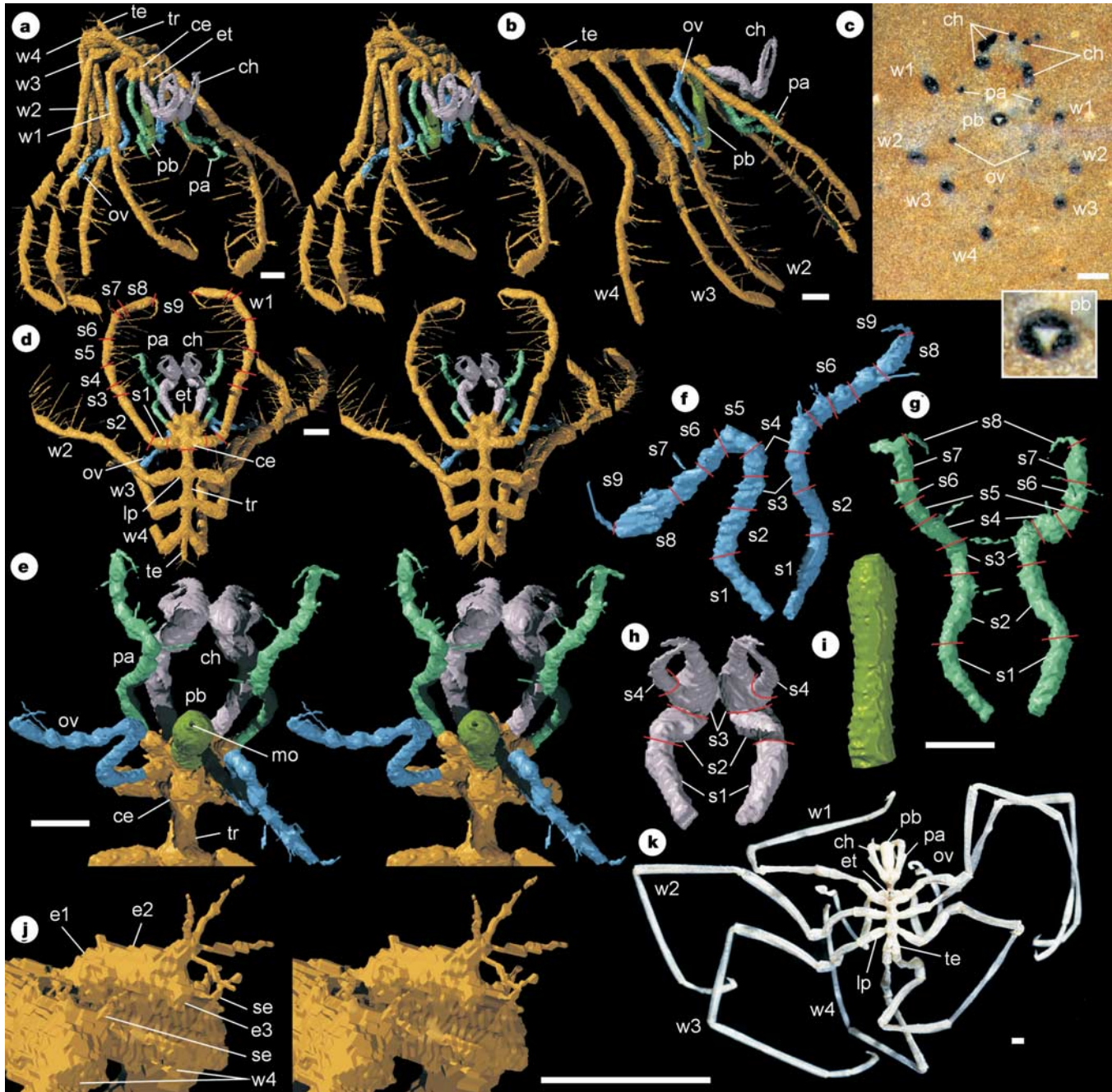


Figure 1 *Haliestes dasos* and the extant pycnogonid *Nymphon gracile*. **a–j**, OUM C.29571, holotype of *Haliestes dasos*. **a, b, d–j**, 'Virtual' reconstructions. **a**, Left anterior oblique stereo pair; **b**, right lateral view; **c**, horizontal section number 67 from serial grinding, with inset of proboscis; **d**, dorsal stereo pair; **e**, cephalosoma and anterior part of trunk, ventral stereo pair, walking legs largely removed; **f**, ovigers, antero-ventral view (scale as in **i**); **g**, palps, antero-dorsal view (scale as in **i**); **h**, chelicerae, dorsal view (scale as in **i**); **i**, proboscis, anterior view (ventral up); **j**, left posterior oblique stereo pair; **k**, dorsal view, *Nymphon gracile*, Recent, Galway Bay, Ireland (Oxford University Museum of Natural

History, Zoological Collections 2003-13-0015). All scale bars 500 μ m. ce, cephalosoma; ch, chelicera; e1–e3, elements of trunk end; et, eye tubercle; lp, lateral process; mo, mouth; ov, oviger; pa, palp; pb, proboscis; s1–s9, segments of appendages; se, setae; te, trunk end; tr, trunk; w1–w4, first to fourth walking legs. Note: w4 distally, on both sides, is not preserved, and the incompleteness, distally, of w2 and w3, right side, and w2, left side, represents data lost in processing. The point of colour change shown here between different structures, such as cephalosoma and chelicerae, is somewhat arbitrary.

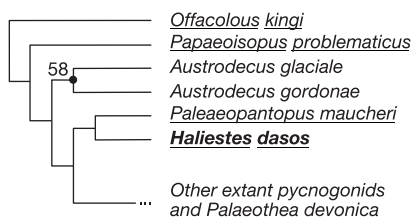


Figure 2 Strict consensus (consistency index = 0.38, retention index = 0.66) of three most-parsimonious trees of length 203 obtained from analysis of data matrix (see Supplementary Table). Bootstrap values, where over 50, are shown. Underlined species are fossils. Most extant species and *Palaeothea devonica* have been removed for clarity (see Supplementary Figure for full tree).

discovered among the diverse arachnomorph arthropods of the Cambrian period to undermine the significance of the chelicerae in this regard. Placement of the pycnogonids as sister to all other euarthropods⁸ receives no new support from this discovery.

The phylogenetic position of *Haliestes* within the pycnogonids was analysed by adding it to a modified version (see Supplementary Table) of the most extensive published character matrix for pycnogonids, which included species from all extant families together with *Palaeoisopus*²⁰. *Palaeopantopus*, *Palaeothea* and the chelicerate *Offacolus*¹² were also added, the last as a non-pycnogonid outgroup. Both *Haliestes* and *Palaeothea* are each based on a single specimen of unknown gender, and were coded on the arbitrary assumption that they are not sexually dimorphic (see Supplementary Methods).

Analysis of this data set using unweighted characters (Fig. 2; Supplementary Figure) suggests that *Haliestes* and *Palaeopantopus* may form a clade within the crown group, with *Palaeoisopus* lying in the stem group. The relative positions of *Palaeoisopus*, *Austrodecus* and *Palaeopantopus* are identical in unfigured implied weights analyses of this revised data set (following the methodology of ref. 20, Fig. 2; see also Supplementary Methods), although here *Haliestes* is in the stem, opposed to the crown group. The basal position of *Austrodecus* was not recovered before the addition of fossil taxa, but concurs with a recent study combining morphological and molecular data²¹. The crown-group position for *Palaeopantopus* differs from previous interpretations^{4,17}. *Palaeothea* falls in a relatively derived position in both analyses, although 24 of the 39 characters for this taxon are coded as missing data. The analysis of Fig. 2 is poorly supported and the phylogenetics of the sea spiders remains to be resolved in detail^{20,21}, hence these results should be considered preliminary—nonetheless, our analyses suggest that *Haliestes* belongs near or in the pycnogonid crown group, which may thus have originated by the Silurian period.

Haliestes inhabited the outer shelf/upper slope of the sub-tropically positioned Anglo-Welsh Basin, and its mode of life and functional morphology were probably like those of extant pycnogonids (Supplementary Note 2). □

Methods

Morphology was reconstructed digitally following serial grinding at 20-μm intervals¹². See Supplementary Methods for details of cladistic analyses.

Received 11 June; accepted 12 August 2004; doi:10.1038/nature02928.

- King, P. E. *Pycnogonids* (Hutchinson, London, 1973).
- Arnaud, F. & Bamber, R. N. The biology of Pycnogonida. *Adv. Mar. Biol.* **24**, 1–96 (1987).
- Child, C. A. The marine fauna of New Zealand: Pycnogonida (sea spiders). *Nat. Inst. Water Atmos. Res. Biodiv. Memoir* **109**, 1–71 (1998).
- Waloszek, D. & Dunlop, J. A. A larval sea spider (Arthropoda: Pycnogonida) from the Upper Cambrian 'Orsten' of Sweden, and the phylogenetic position of pycnogonids. *Palaeontology* **45**, 421–446 (2002).
- Vilpoux, K. & Waloszek, D. Larval development and morphogenesis of the sea spider *Pycnogonum littorale* (Ström, 1762) and the tagmosis of the body of Pantopoda. *Arthrop. Struct. Dev.* **32**, 349–383 (2003).
- Boxshall, G. A. The evolution of arthropod limbs. *Biol. Rev.* **79**, 253–300 (2004).
- Dunlop, J. A. & Arango, C. P. Pycnogonid affinities: a review. *J. Zool. Syst. Evol. Res.* (in the press).
- Giribet, G., Edgecombe, G. D. & Wheeler, W. C. Arthropod phylogeny based on eight molecular loci and morphology. *Nature* **413**, 157–161 (2001).

- Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Soft-bodied fossils from a Silurian volcanoclastic deposit. *Nature* **382**, 248–250 (1996).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. An exceptionally preserved vermiform mollusc from the Silurian of England. *Nature* **410**, 461–463 (2001).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. A three-dimensionally preserved fossil polychaete worm from the Silurian of Herefordshire, England. *Proc. R. Soc. Lond. B* **268**, 2355–2363 (2001).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J., Siveter, Derek J. & Orr, P. J. The arthropod *Offacolus kingi* (Chelicerata) from the Silurian of Herefordshire, England: computer based morphological reconstructions and phylogenetic affinities. *Proc. R. Soc. Lond. B* **269**, 1195–1203 (2002).
- Siveter, David J., Sutton, M. D., Briggs, D. E. G. & Siveter, Derek J. An ostracode crustacean with soft parts from the Lower Silurian. *Science* **302**, 1749–1751 (2003).
- Briggs, D. E. G., Sutton, M. D., Siveter, David J. & Siveter, Derek J. A new phyllocarid (Crustacea: Malacostraca) from the Silurian Fossil-Lagerstätte of Herefordshire, UK. *Proc. R. Soc. Lond. B* **271**, 131–138 (2004).
- Orr, P. J., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Three-dimensional preservation of a non-biomineralized arthropod in concretions in Silurian volcanoclastic rocks from Herefordshire, England. *J. Geol. Soc. Lond.* **157**, 173–186 (2000).
- Miyazaki, K. J. On the shape of the foregut lumen in sea spiders (Arthropoda: Pycnogonida). *Mar. Biol. Ass. UK* **82**, 1037–1038 (2002).
- Bergström, J., Stürmer, W. & Winter, G. *Palaeoisopus*, *Palaeopantopus* and *Palaeothea*, pycnogonid arthropods from the Lower Devonian Hunsrück Slate, West Germany. *Paläont. Zeit.* **54**, 7–54 (1980).
- Bartels, C., Briggs, D. E. G. & Bräse, G. *The Fossils of the Hunsrück Slate* (Cambridge Univ. Press, 1998).
- Wilby, P. R., Briggs, D. E. G. & Riou, B. Mineralization of soft-bodied invertebrates in a Jurassic metalliferous deposit. *Geology* **24**, 847–850 (1996).
- Arango, C. P. Morphological phylogenetics of the sea spiders (Arthropoda: Pycnogonida). *Org. Divers. Evol.* **2**, 107–125 (2002).
- Arango, C. P. Molecular approach to the phylogenetics of sea spiders (Arthropoda: Pycnogonida) using partial sequences of nuclear ribosomal DNA. *Mol. Phylogenet. Evol.* **28**, 588–600 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by the Leverhulme Trust, NERC and English Nature. K. Saunders is thanked for technical assistance; C. Arango, R. Bamber and D. Waloszek for discussion; G. Boxshall, J. Dunlop and D. Waloszek for pre-prints of their papers; S. de Grave for providing Recent material; and R. Fenn, T. Hall and J. Sinclair for general assistance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.S. (derek.siveter@earth.ox.ac.uk).

Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous

Bryan Kolaczkowski¹ & Joseph W. Thornton²

¹Department of Computer and Information Science and ²Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97403, USA

All inferences in comparative biology depend on accurate estimates of evolutionary relationships. Recent phylogenetic analyses have turned away from maximum parsimony towards the probabilistic techniques of maximum likelihood and bayesian Markov chain Monte Carlo (BMCMC). These probabilistic techniques represent a parametric approach to statistical phylogenetics, because their criterion for evaluating a topology—the probability of the data, given the tree—is calculated with reference to an explicit evolutionary model from which the data are assumed to be identically distributed. Maximum parsimony can be considered nonparametric, because trees are evaluated on the basis of a general metric—the minimum number of character state changes required to generate the data on a given tree—without assuming a specific distribution¹. The shift to parametric methods was spurred, in large part, by studies showing that although both approaches perform well most of the time², maximum parsimony is strongly biased towards recovering an

incorrect tree under certain combinations of branch lengths, whereas maximum likelihood is not^{3–6}. All these evaluations simulated sequences by a largely homogeneous evolutionary process in which data are identically distributed. There is ample evidence, however, that real-world gene sequences evolve heterogeneously and are not identically distributed^{7–16}. Here we show that maximum likelihood and BMCMC can become strongly biased and statistically inconsistent when the rates at which sequence sites evolve change non-identically over time. Maximum parsimony performs substantially better than current parametric methods over a wide range of conditions tested, including moderate heterogeneity and phylogenetic problems not normally considered difficult.

Functional constraints on sites in a gene sequence often change through time, causing shifts in site-specific evolutionary rates, a phenomenon called heterotachy (meaning ‘different speeds’)^{7–16}. When an identically distributed evolutionary framework is imposed on sequences that evolve heterogeneously, parameter estimates are compromised over sites and lineages and are therefore incorrect for many or all sites. Likelihood-based techniques are guaranteed to recover the true phylogeny only when the correct model is used, and nonparametric statistical methods are often applied when the assumptions of parametric techniques are violated. On the other hand, parametric methods, including maximum likelihood, are generally more powerful than nonparametric techniques and can be robust to certain violations^{17,18}. We used an experimental approach to evaluate the phylogenetic accuracy of parametric and nonparametric methods under a simple form of heterotachy. We simulated replicate DNA sequence alignments with two symmetrical rate partitions along a four-taxon tree; each partition represents

a phylogenetically challenging problem—two clades, each consisting of a long branch (length p) and a short branch (length q)—but the sites with accelerated rates differ between partitions (Fig. 1a). To reveal the specific impact of heterogeneity, we compared phylogenetic accuracy (the fraction of replicates from which the true tree was recovered) on heterogeneous data with accuracy on control sequences simulated under corresponding evolutionary conditions without heterogeneity (see Methods).

Under conditions of strong heterotachy ($p = 0.75$ substitutions per site, $q = 0.05$), the accuracy of both maximum likelihood and BMCMC is dramatically reduced compared with homogeneous controls (Fig. 1b). Both methods have zero accuracy when the internal branch length $r < 0.22$, and they reach 100% accuracy only when $r > 0.34$. Maximum parsimony is superior to the parametric methods when $0.15 < r < 0.35$, and it is never inferior. For each method, we used nonlinear regression to estimate the internal branch length at which 50% accuracy is achieved (BL_{50}) and found that maximum parsimony can reliably recover the true topology at significantly shorter internal branch lengths ($BL_{50} = 0.22$) than the two likelihood-based methods ($BL_{50} = 0.28$, $P < 0.001$). Maximum parsimony's performance is worse than that of the parametric methods on single-partition data (due to the well-known long branch attraction bias⁵), but it is not additionally hampered by evolutionary heterogeneity ($P = 0.76$). Maximum parsimony retains its performance advantage over maximum likelihood and BMCMC on heterotachous data when strong support is required to accept a tree as resolved (bootstrap or posterior probability $> 95\%$, Supplementary Fig. S1). These results indicate that heterotachy substantially reduces the accuracy of maximum likelihood and BMCMC on phylogenetic problems

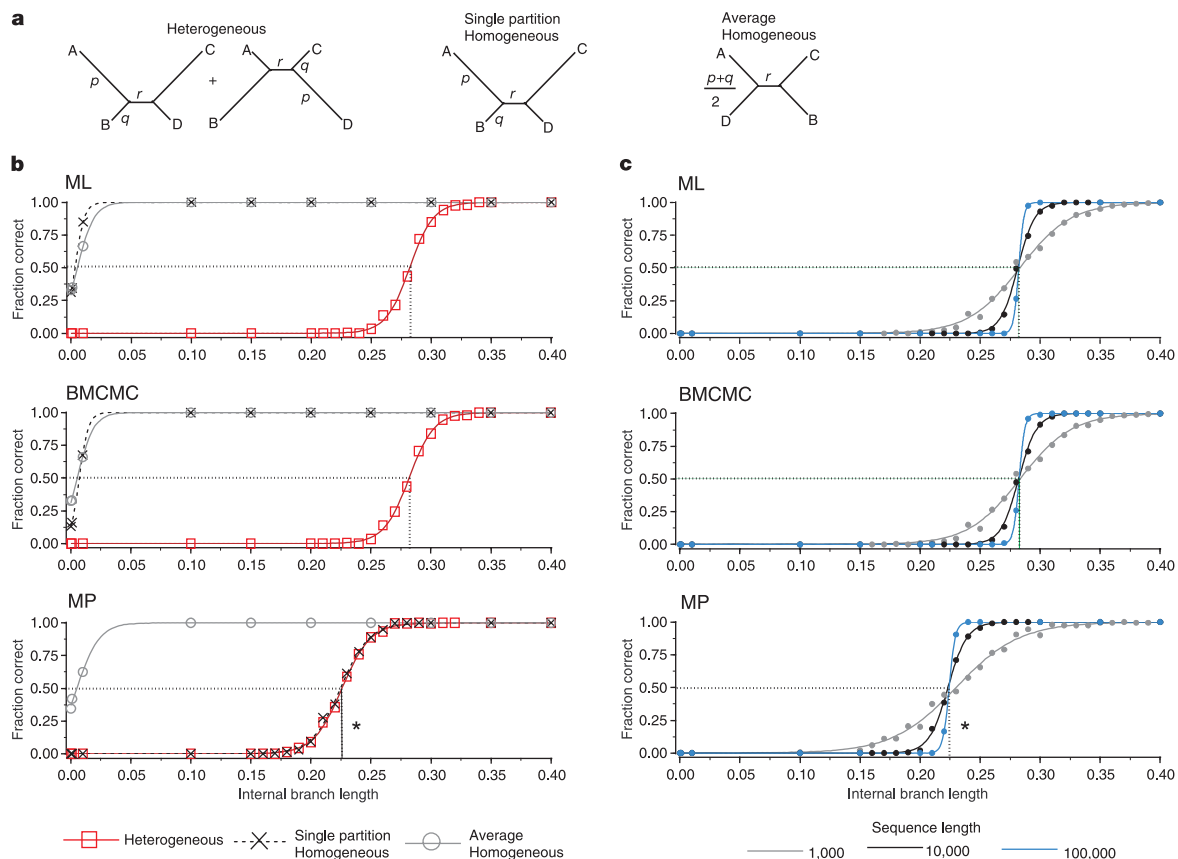


Figure 1 Likelihood-based methods are less accurate than maximum parsimony (MP) under heterogeneous conditions. **a**, Trees on which heterogeneous and control sequences were simulated. **b**, Heterotachy reduces the accuracy of likelihood methods. Accuracy is plotted against internal branch length for sequences with and without strong

heterotachy. Dotted lines, BL_{50} for each method (asterisk: maximum parsimony $<$ maximum likelihood (ML) and BMCMC, $P < 0.001$). **c**, Likelihood methods are inconsistent below the BL_{50} under strong heterotachy, recovering the incorrect tree with increasing frequency as the amount of data increases.

that are not difficult enough to impair maximum parsimony.

Under the heterotachous conditions studied, maximum likelihood and BMCMC are statistically inconsistent, converging on the wrong answer as the amount of data grows. For internal branch lengths below the BL_{50} , accuracy declines to zero as sequence length increases, indicating that parametric methods are statistically inconsistent in this region of parameter space; the BL_{50} therefore represents an 'inconsistency point' (Fig. 1c and Supplementary Fig. S2). This inconsistency is due to a directional bias: maximum likelihood and BMCMC specifically infer the erroneous tree ((AC),(BD)) with high support when the internal branch is shorter than the BL_{50} , including length zero (Supplementary Fig. S3). This is the same tree towards which maximum parsimony is biased on single-partition data, but heterotachy causes likelihood-based methods to infer the incorrect tree over a wider range of parameter values and with stronger apparent support.

Heterotachy reduces the performance of parametric methods across a broad range of evolutionary conditions. Whenever the short terminal branch length $q < 0.3$, maximum parsimony significantly outperforms both likelihood-based methods. Even fairly weak heterotachy—a ratio of branch lengths among partitions as low as 0.5:0.2—is sufficient to produce a significant performance disparity between the likelihood-based methods and maximum parsimony (Fig. 2a). The more intense the heterotachy, the greater the performance difference. Furthermore, maximum likelihood's accuracy can be reduced to zero when only a small fraction of sites deviate in rate from the rest of the sequence. Fewer heterotachous sites are required to impair performance as heterotachy grows more intense or the phylogenetic problem becomes more difficult (Fig. 2b).

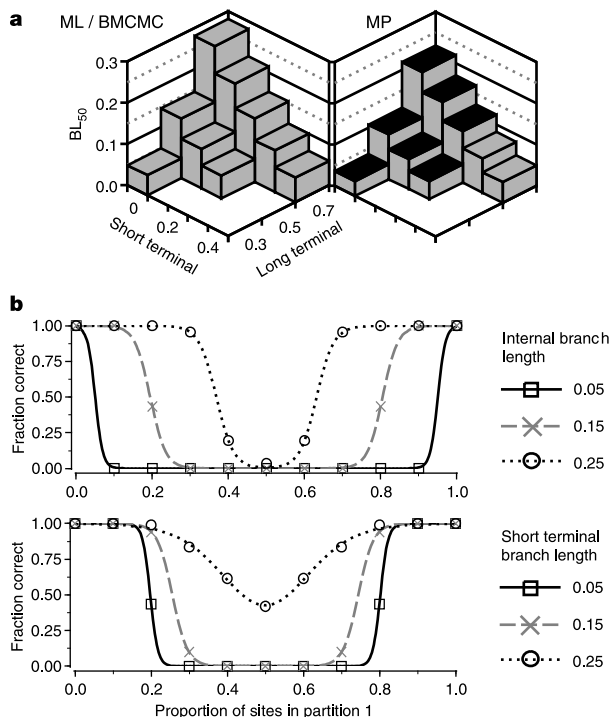


Figure 2 Parsimony outperforms likelihood over a wide range of heterotachous conditions. **a**, Maximum parsimony is more accurate than likelihood-based methods on data with weaker heterotachy. Bars show the BL_{50} for combinations of long and short terminal branch lengths in heterotachous data sets (black: maximum parsimony < maximum likelihood and BMCMC, $P < 0.001$). The BL_{50} s for maximum likelihood and BMCMC are equivalent for all conditions ($P > 0.91$). **b**, Maximum likelihood accuracy is impaired when only a small fraction of sites are heterotachous. Accuracy is plotted against the fraction of heterotachous sites as the phylogenetic problem becomes more difficult (upper panel: $p = 0.75$, $q = 0.05$) and heterotachy more intense (lower panel: $p = 0.75$, $r = 0.15$).

We used several existing likelihood models that account for among-site or -lineage rate variation by applying identically distributed models of heterogeneity, including gamma, invariant sites and covarion models, but none improve the performance of maximum likelihood or BMCMC (Fig. 3a). Using amino acid instead of nucleotide sequences substantially increases the accuracy of maximum parsimony ($BL_{50} = 0.08$) because convergence is less likely with 20 than with 4 possible states. In contrast, maximum likelihood and BMCMC improve to a much smaller extent ($BL_{50} = 0.18$). As a result, maximum parsimony's performance advantage on heterotachous protein sequences is even greater than on DNA (Fig. 3b). Denser taxon sampling to break up long branches¹⁹ improves the accuracy of all methods by about equal proportions (Fig. 3b).

The accuracy of likelihood-based methods declines because they erroneously impose homogeneous branch lengths across sites. On heterotachous data with internal branch lengths below the inconsistency point, maximum likelihood overestimates the length of the internal branch and infers the lengths of the long and short terminals as approximately the average over the two partitions (Supplementary Fig. S4). To test whether these errors are responsible for phylogenetic bias, we compared the standard homogeneous maximum likelihood model (ML_{homo}) with an a priori partitioned model in which the branch lengths for each site are constrained to their true values (ML_{true}). As Fig. 4a shows, ML_{true} has much better performance ($P < 0.001$). Models that set only the internal (ML_{term}) or the internal and long terminal branches (ML_{short}) to their true lengths did not improve performance. Correcting the short terminal (ML_{long}), however, yields a substantial improvement in phylogenetic accuracy. Erroneous optimization of the short terminal length using 'compromise' branch lengths is therefore the primary cause of heterotachy-induced phylogenetic error in maximum likelihood (Fig. 4a).

Maximum likelihood's bias is caused by misinterpretation of specific character state patterns. We analysed the contribution each character state pattern makes to the likelihood of the true and erroneous trees and compared net support for the true tree using ML_{homo} to that using the heterogeneous model ML_{true} (Fig. 4b). Patterns that provide the most support for the correct tree under ML_{true} ($xxxy$ and $xyxz$) only weakly support the true tree when ML_{homo} is used; this occurs because ML_{homo} overestimates the probability that these patterns are due to convergence on branches whose lengths are overestimated. In contrast, the convergent patterns $xyxy$ and $xyxz$ support the wrong tree using either method. As a result, the likelihood of the incorrect tree becomes greater than that of the true tree when ML_{homo} is used on heterotachous data. Under the same conditions, maximum parsimony recovers the true

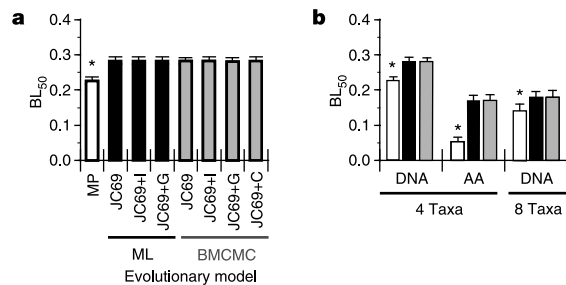


Figure 3 Maximum parsimony is more accurate than likelihood methods when techniques to improve phylogenetic performance are used. **a**, Accuracy of likelihood-based methods on heterotachous data does not improve when evolutionary models that incorporate among-site rate variation (+G, gamma distribution; +I, invariant sites) or covarion heterotachy (+C) are used. BL_{50} s are shown under strong heterotachy; bars indicate 99% confidence intervals. Asterisks show lower BL_{50} values ($P < 0.001$). **b**, Maximum parsimony (white) outperforms maximum likelihood (black) and BMCMC (grey) on amino acid sequences and 8-taxon data sets with strong heterotachy.

tree because the frequency of $xyxy$ is greater than that of $xyxz$; the patterns $xyxz$ and $xyxz$, which taken together mislead ML_{homo} , are not informative in a nonparametric context.

The bias of parametric methods arises due to heterogeneity in the data and the resulting violation of the identical distribution assumption, as predicted theoretically²⁰. We implemented a novel likelihood method using a mixed model ($BMCMC_{\text{hetero}}$) that incorporates heterotachy by including two branch length sets for each topology. For each sequence site, the likelihood is calculated for each branch length set, weighted by the posterior probability of the site being in that set and then summed to yield the total likelihood. This model, which corresponds to the true evolutionary conditions but assuming an identical data distribution, performs dramatically better than both maximum parsimony and the standard maximum likelihood or $BMCMC$ algorithms ($BL_{50} = 0.045$, $P < 0.001$) on heterotachous data (Fig. 4c). It did not perform as well, however, as a non-identically distributed method ($BMCMC_{\text{true}}$) that uses the true evolutionary model with a priori sorting of sites into their true partitions. Furthermore, $BMCMC_{\text{hetero}}$ remains statistically inconsistent, converging on the wrong tree as sequence length increases at internal branch lengths $r < BL_{50}$, (Supplementary Fig. S5). $BMCMC_{\text{true}}$ is consistent under all conditions examined. These results indicate that violating the identical distribution assumption can cause inconsistency, even when the 'true' evolutionary model is used.

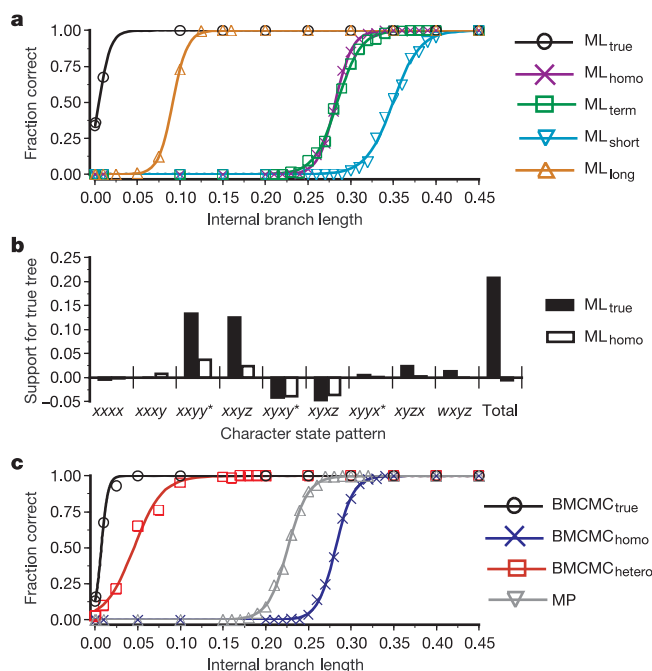


Figure 4 Poor maximum likelihood performance is due to assuming homogeneous branch lengths. **a**, Maximum likelihood error is caused primarily by overestimating short terminal branch lengths due to heterogeneity. Accuracy on strongly heterotachous sequences is shown as the internal branch length increases, using several likelihood models that constrain all (ML_{true}), some (ML_{term} , ML_{short} , ML_{long}) or no branches on the tree (ML_{homo}) to their true lengths for all sites. **b**, Support for the true tree by specific character state patterns is reduced due to strong heterogeneity when ML_{homo} is used. For each character state pattern and model, net support is shown as the ratio of the likelihood of the true topology to the likelihood of the incorrect ((AC),(BD)) tree, weighted by the frequency of the pattern. Asterisks indicate parsimony-informative patterns. **c**, Incorporating heterotachy improves the accuracy of parametric methods. Accuracy on strongly heterotachous data are shown for the homogeneous model ($BMCMC_{\text{homo}}$), a model that allows two independent branch length sets and correct a priori partitioning of sites ($BMCMC_{\text{true}}$), and a novel model with two branch length sets and likelihoods calculated on the basis of a posteriori weighting ($BMCMC_{\text{hetero}}$).

The form of heterotachy studied here is only one way that heterotachy can be distributed on a tree. Our additional work (not shown) indicates that several other forms of heterotachy can also impair the accuracy of parametric methods. The evolutionary model used in our simulations is a simplified one; the extent to which phylogenetic accuracy is impaired by the more complex evolutionary dynamics likely to affect real-world sequences is currently unknown. There are numerous sequence data sets from which parametric methods have failed to infer otherwise well-corroborated phylogenies^{21–24}, including one in which heterotachy has recently been implicated¹³.

There are two ways to avert the negative effects of heterogeneity on parametric methods. One is to use maximum parsimony, which is not affected by heterotachy because it does not assume an identically distributed evolutionary process. The other is to develop more complex parametric models. Our results indicate that a new likelihood method using mixed branch length models may offer substantially improved accuracy on heterotachous sequences, but there are reasons for caution. The model that performed well in our tests matched the true evolutionary process, which we knew a priori. With real sequences, we do not know the true number of branch length partitions, so imposed models will usually use either too many or too few branch length parameters. For many sequences, the actual number of branch length categories may approach the number of sites; under these conditions, the true one-category-per-site likelihood model is formally equivalent to maximum parsimony²⁵. Finally, the computational burden of mixed-model phylogenetic inference grows exponentially with the number of branch length sets. With current algorithms and computing power, incorporating heterotachy into a likelihood framework will often require sacrifices in the number of sequences analysed or the rigor with which tree and parameter space are searched, which may also reduce phylogenetic accuracy¹.

Our findings place those who infer and use phylogenetic trees in an uncertain position. Previous research has shown that parametric methods are superior or equal to nonparametric approaches when evolutionary heterogeneity is not present, but our work shows that maximum parsimony can substantially outperform current likelihood-based methods when it is. Worse still, heterotachy-induced bias leaves no obvious signature because the inferred trees have moderate branch lengths and strong support for erroneous nodes. With no reliable a posteriori diagnostic for heterotachy-induced phylogenetic error, how can we know which method to choose or, when trees from different methods conflict, which one to favour? The overall frequency and severity of the conditions that favour likelihood as compared with those that favour parsimony is not yet known for real-world sequences. At present, we recommend reporting nonparametric analyses along with parametric results and interpreting likelihood-based inferences with the same caution now applied to maximum parsimony trees. In the future, it is possible that new mixed-model techniques may improve likelihood's performance to the point that it is consistently superior to nonparametric methods. □

Methods

Simulations

We simulated sequences along a 4-taxon tree ((A,B),(C,D)) with two independent partitions that were concatenated into one heterogeneous alignment. In one partition, long terminal branches ($p \in (0.3, 0.75)$) lead to A and C, and short terminals ($q \in (0.001, 0.4)$) lead to B and D. In the other partition, terminal branches to B and D have length p , whereas A and C have length q . The internal branch length ($r \in (0.0, 0.5)$) is equal in both partitions. The two partitions were of equal size unless otherwise noted. Two-hundred replicate alignments of 1,000, 5,000, 10,000 and 100,000 characters were simulated under each set of conditions using the JC69 (DNA) or Poisson (protein) model. Average homogeneous control data were simulated using the same internal branch length as in the experimental condition and terminals with the mean length over the two partitions. Single-partition homogeneous controls were simulated using conditions for one of the experimental partitions (Fig. 1a). Sequences were also simulated on 8-taxon trees derived from 4-taxon trees by bisecting each terminal branch at the halfway point.

Phylogenetic analysis

Phylogenies were analysed using PAUP* v4.0b10 (ref. 26), PAML v3.14 (ref. 27) and MrBayes v3.0b4 (ref. 28). To determine the best-fit likelihood model for nucleotide data, hierarchical likelihood ratio tests were performed on 100 randomly selected alignments chosen from experimental data sets (using Modeltest v3.06 (ref. 29), $\alpha = 0.05$). The true JC69 model was strongly supported and was used for all maximum likelihood and BMCMC DNA analyses. The true Poisson model was used for protein analysis, and maximum parsimony used equal weights. We performed likelihood-based analyses with and without gamma and invariant sites models to determine their effect on accuracy. The covarion model implemented in MrBayes was also used. Topology searches were exhaustive for maximum parsimony and maximum likelihood. BMCMC analysis involved four chains (three heated) run well past stationarity.

To determine support, we used nonparametric bootstrapping (1,000 replicates) for maximum parsimony and maximum likelihood and posterior probability for BMCMC, with a support cutoff value of 95% to construct strongly supported consensus trees. (See Supplementary Information for details on phylogenetic methods.)

Accuracy

The accuracy of each method was calculated as the proportion of replicates for which the correct topology was uniquely recovered (ϕ). Nonlinear regression was performed using the logistic equation $\phi = 1/(1 + \exp((BL_{50} - r)H))$, in which BL_{50} is the estimated internal branch length that produces 50% correct recovery, and H estimates the steepness of the performance curve. The significance of differences among BL_{50} s was examined by a t -test.

Bias and error

The type I error rate for each method was determined by analysing data sets generated under strong heterotachy with zero-length internal branches and determining the fraction of replicates falsely resolved with 95% bootstrap or posterior probability support³⁰. The presence of bias was determined by calculating the proportion of erroneous estimates consistent with each possible incorrect topology over all internal branch lengths. The intensity of bias was investigated by calculating the proportion of erroneous topology estimates consistent with each possible incorrect topology when a 95% support cutoff was imposed.

To determine the impact of homogeneous optimization of branch lengths on maximum likelihood error, we compared the standard maximum likelihood algorithm that estimates a single set of branch lengths (ML_{homo}) with several partitioned maximum likelihood models with constrained branch lengths. ML_{true} constrains all branch lengths for each site to the true values used to simulate data sets. ML_{term} constrains the internal branch lengths to the true value for each site, but terminal branches have the lengths homogeneously optimized under ML_{homo} . ML_{short} assumes the true internal and long terminal branches but uses the short terminal length from ML_{homo} . ML_{long} constrains the internal and short terminal branches to their true values and takes the long terminal branch length from ML_{homo} .

Support for the true topology by each character-state pattern was calculated from a 100,000-site data set constructed under strong heterotachy ($p = 0.75$, $q = 0.05$, $r = 0.254$). Net support for the true tree is defined as the likelihood ratio of the true tree to the incorrect tree for each pattern x , weighted by the frequency of x ($f(x)$) in the data set: $S_{((AB),(CD)),x} = \frac{P(x|((AB),(CD)))}{P(x|((AC),(BD)))} f(x)$.

To determine the performance impact of violating the identical distribution assumption when the true evolutionary model is used, we implemented a novel likelihood model ($BMCMC_{\text{hetero}}$) that incorporates heterotachy a posteriori by applying two sets of branch lengths to the data. For each sequence site x_i , the likelihood of tree t with branch length sets b_1 and b_2 is $L(t|x_i) = \sum_{j=1}^2 [\rho_{ij} P(x_i|t, b_j)]$, where ρ_{ij} —the posterior probability that x_i is in branch length set b_j —is calculated from the data as $\rho_{ij} = P(x_i|t, b_j) / \sum_{k=1}^2 P(x_i|t, b_k)$. The overall posterior probability of each topology is calculated using BMCMC (see Supplementary Information). This new method was compared with a BMCMC analysis using the true heterogeneous model and correct a priori data partitioning ($BMCMC_{\text{true}}$).

Received 17 May; accepted 6 August 2004; doi:10.1038/nature02917.

- Sanderson, M. J. & Kim, J. Parametric phylogenetics? *Syst. Biol.* **49**, 817–829 (2000).
- Hillis, D. M., Huelsenbeck, J. P. & Cunningham, C. W. Application and accuracy of molecular phylogenies. *Science* **264**, 671–677 (1994).
- Felsenstein, J. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
- Kuhner, M. K. & Felsenstein, J. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11**, 459–468 (1994).
- Huelsenbeck, J. P. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biol.* **47**, 519–537 (1998).
- Gaut, B. S. & Lewis, P. O. Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.* **12**, 152–162 (1995).
- Huelsenbeck, J. P. Testing a covariotide model of DNA substitution. *Mol. Biol. Evol.* **19**, 698–707 (2002).
- Miyamoto, M. M. & Fitch, W. M. Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol.* **12**, 503–513 (1995).
- Lopez, P., Casane, D. & Philippe, H. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* **19**, 1–7 (2002).
- Fitch, W. M. The molecular evolution of cytochrome c in eukaryotes. *J. Mol. Evol.* **8**, 13–40 (1976).
- Pollock, D. D., Taylor, W. R. & Goldman, N. Coevolving protein residues: maximum likelihood identification and relationship to structure. *J. Mol. Biol.* **287**, 187–198 (1999).
- Pupko, T. & Galtier, N. A covarion-based method for detecting molecular adaptation: application to the evolution of primate mitochondrial genomes. *Proc. R. Soc. Lond. B* **269**, 1313–1316 (2002).
- Inagaki, Y., Susko, E., Fast, N. M. & Roger, A. J. Covarion shifts cause a long-branch attraction artifact that unites Microsporidia and Archaeobacteria in EF-1[alpha] phylogenies. *Mol. Biol. Evol.* **21**, 1340–1349 (2004).

- Lockhart, P. J. *et al.* A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol.* **15**, 1183–1188 (1998).
- Misof, B. *et al.* An empirical analysis of mt 16S rRNA covarion-like evolution in insects: site-specific rate variation is clustered and frequently detected. *J. Mol. Evol.* **55**, 460–469 (2002).
- Philippe, H. & Lopez, P. On the conservation of protein sequences in evolution. *Trends Biochem. Sci.* **26**, 414–416 (2001).
- Donaldson, T. S. Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. *J. Am. Stat. Assoc.* **63**, 660–676 (1968).
- Sullivan, J. & Swofford, D. L. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Syst. Biol.* **50**, 723–729 (2001).
- Hillis, D. M. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* **47**, 3–8 (1998).
- Chang, J. T. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* **134**, 189–215 (1996).
- Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804 (2003).
- Russo, C. A., Takezaki, N. & Nei, M. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* **13**, 525–536 (1996).
- Delarbre, C., Gallut, C., Barriel, V., Janvier, P. & Gachelin, G. Complete mitochondrial DNA of the hagfish, *Eptatretus burgeri*: the comparative analysis of mitochondrial DNA sequences strongly supports the cyclostome monophyly. *Mol. Phylogenet. Evol.* **22**, 184–192 (2002).
- Naylor, G. J. & Brown, W. M. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* **47**, 61–76 (1998).
- Tuffley, C. & Steel, M. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.* **59**, 581–607 (1997).
- Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods*, v4.0b10 (Sinauer Associates, Sunderland, Massachusetts, 1998).
- Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
- Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
- Swofford, D. L. *et al.* Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* **50**, 525–539 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Conery for advice, support and programming advice. P. Phillips, R. DeSalle and S. Proulx provided comments and discussion. We benefited from discussions of mixed model methods with D. Zwickl. B.K. was supported by an NSF IGERT training grant in Evolution, Development and Genomics to the University of Oregon.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T. (joet@uoregon.edu).

Ecological constraints on diversification in a model adaptive radiation

Rees Kassen^{1,2}, Martin Llewellyn^{1,*} & Paul B. Rainey^{1,3}

¹Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

²Department of Biology & Centre for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa K1N 6N5, Canada

³School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand

* Present address: London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Taxonomic diversification commonly occurs through adaptive radiation, the rapid evolution of a single lineage into a range of genotypes or species each adapted to a different ecological niche^{1,2}. Radiation size (measured as the number of new types) varies widely between phylogenetically distinct taxa^{2–4} and between replicate radiations within a single taxon where the ecological opportunities available seem to be identical^{5,6}. Here we

show how variation in energy input (productivity) and environmental disturbance combine to determine the extent of diversification in a single radiating lineage of *Pseudomonas fluorescens* adapting to laboratory conditions. Diversity peaked at intermediate rates of both productivity and disturbance and declined towards the extremes in a manner reminiscent of well-known ecological patterns^{7–9}. The mechanism responsible for the decrease in diversity arises from pleiotropic fitness costs associated with niche specialization^{10,11}, the effects of which are modulated by gradients of productivity and disturbance. Our results indicate that ecological gradients may constrain the size of adaptive radiations, even in the presence of the strong diversifying selection associated with ecological opportunity, by decoupling evolutionary diversification from ecological coexistence.

The size of an adaptive radiation is thought to be governed by the extent of ecological opportunity (defined as the number of vacant niches or unutilized resources; where ‘niches’ refers to the collection of sites within an environment offering different conditions of growth) in a given geographical region or community^{1,2,12}. This requires the processes of evolutionary diversification (which introduces variant types into the community) and ecological assortment (which can eliminate new types) to be closely matched. Any factor that influences the relative rates at which these two processes occur is likely to have an impact on the size of a given adaptive radiation such that diversification may be less (or more) extensive than predicted on the basis of ecological opportunity alone. Relevant factors include contingent or idiosyncratic increases in the probability of diversification or extinction through sexual selection¹³, drift⁷ or loss of habitat associated with decreasing area¹⁴. A more general factor may arise from pleiotropic fitness costs associated with niche specialization, particularly if the magnitude of the cost is affected by ecological conditions^{2,3,15}.

To understand the effect of ecological constraints on diversification we allowed replicate populations of a single genotype of the soil bacterium *P. fluorescens* SBW25 to diversify in spatially structured (static) microcosms across gradients of productivity and disturbance over the course of 16 days in the absence of predation. Ecological opportunity is afforded by the combination of spatial structure of the microcosms and resource depletion in the broth caused by the growth and metabolism of the ancestral population. Under these conditions, *P. fluorescens* rapidly diversifies into niche specialist genotypes that differ in colony morphology on agar plates and occupy different ecological regions of a static microcosm¹¹. The diversity in these communities is non-neutral, being stably maintained by negative frequency-dependent selection, and the extent of diversification has been shown to be related to nutrient concentration¹⁶. When spatial structure is removed by shaking, diversification of the kind observed in the presence of spatial structure does not occur and diversity is rapidly lost from previously diverse cultures^{11,17,18}.

Figures 1 and 2 depict the response of diversity and richness (number of morphotypes) to gradients of both productivity and disturbance after 16 days. There is a peak at intermediate rates of both, a result confirmed by an analysis of covariance showing a significant negative quadratic relationship between diversity and both productivity and disturbance but no significant linear relationship (Table 1). This bivariate pattern of diversity is also consistent with purely ecological models of competition between niche specialists under resource competition^{8,19–21} and in the presence of coarse-grained spatial heterogeneity^{17,18,22,23} (see Supplementary Information). Note, however, that as interpretations of our data these models are not appropriate because they do not account for diversification from an initially isogenic state.

The two dominant niche specialists in our experiment were the ancestral smooth, which occupies the broth, and wrinkly spreader, which colonizes the air–broth interface. Note that a single smooth

morphotype dominates at the extremes of both productivity and disturbance (Fig. 2a, e). We did observe other types (including the fuzzy spreader morph that inhabits the bottom of the microcosm and many variant types within categories), however these never comprised more than 5% of the total population so we do not consider them further.

These results are consistent with previous work, which has highlighted the importance of spatial structure^{11,17,18} and resource competition^{11,16,24} for the emergence and maintenance of diversity; the results are also consistent with our knowledge of the evolutionary ecology of *P. fluorescens* in static microcosms. In spatially structured microcosms the key limiting resource seems to be oxygen. Shallow oxygen gradients arise even in the absence of bacteria, owing to diffusion and the special properties of liquid surfaces. These gradients are steepened by the metabolic activities of growing bacteria, causing intense resource competition that drives diversification. This interpretation is consistent with the ecological theory of adaptive radiation^{2,15} that sees diversification being spurred initially by divergent natural selection underlain by environmental heterogeneity and subsequently driven by resource competition.

Mutation in the ancestral smooth population constantly introduces new variants into the population where, under standard conditions (intermediate nutrient concentrations and disturbance frequencies), strong diversifying selection favours the evolution of niche specialist genotypes—one of the most successful being the wrinkly spreader which, by overproducing an adhesive polymer²⁵, can grow at the air–liquid interface where oxygen is abundant. Diversity is maintained through negative frequency-dependent selection generated through the combination of resource competition and spatial structure. Under frequent disturbances and low nutrient concentrations, population sizes are large enough (approximately 2×10^7 cells ml^{−1}) to guarantee a constant supply of new mutants, including wrinkly spreader, but not to permit formation of a coherent mat, which requires expression of an energetically costly adhesive cellulosic polymer^{10,24,25}. At high nutrient concentrations and infrequent disturbances, intense resource

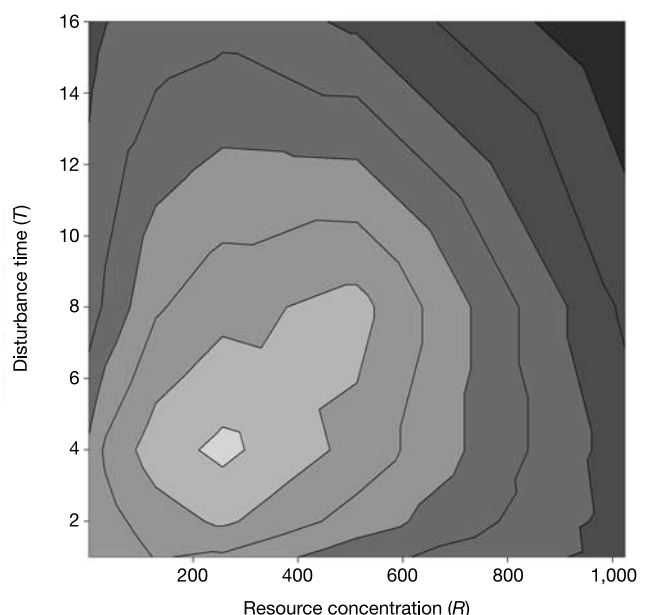


Figure 1 The joint effect of productivity and disturbance on diversity, as $1 - \lambda$, for the experiment. Data were transformed by Loess smoothing using a sampling proportion of 0.5. High diversity regions are shown in lighter shades, low diversity regions in darker shades. Values of diversity may be read directly from the graphs in Fig. 2.

competition leads to strong diversifying selection and the evolution of a thick wrinkly spreader mat. Eventually the mat sinks under its own weight. The collapse is further hastened by smooth morphotypes that act as cheats by invading the mat, thereby gaining access to oxygen, but contributing nothing towards mat strength²⁴.

The overproduction of cellulose characteristic of wrinkly spreader is accompanied by a pleiotropic fitness cost¹⁰, the precise effects of which appear to be dependent on the ecological gradients of productivity and disturbance. Experimental evidence supporting this interpretation is shown in Fig. 3, which depicts the population density of smooth and wrinkly spreader genotypes as pure cultures and in competitive mixtures over 4 days across five levels of productivity. The population density of the smooth morphotypes is typically 1–2 orders of magnitude higher than wrinkly spreader after 1 day of growth in both pure and mixed cultures, reflecting the inability of wrinkly spreader to form a coherent mat in overnight cultures and implying that the smooth morphotype is the superior

competitor under frequent disturbances. On day two and after, the wrinkly spreader morphotype typically achieves population densities comparable to that of smooth in pure culture (Fig. 3a–e) across the entire range of productivities, but is inferior to smooth at low ($0.00781 \times$, $0.125 \times$) and high ($8 \times$) resource supply rates (Fig. 3f, g, j). After 4 days of growth at the highest productivity level ($8 \times$), the density of smooth is moderately but significantly greater when cultured in the presence of wrinkly spreader than in its absence (one-tailed *t*-test: $t = 4.64$; $P = 0.005$, degrees of freedom = 2; compare Fig. 3e, j), whereas that of wrinkly spreader is again lower when competing against smooth (Fig. 3e, j). This result is consistent with the idea that the smooth genotype gains a fitness advantage at the air–broth interface, hastening the collapse of the mat²⁴.

The reduced competitive ability of wrinkly spreader relative to smooth at the extremes of productivity and disturbance could arise through changes in either the relative fitness of either type in

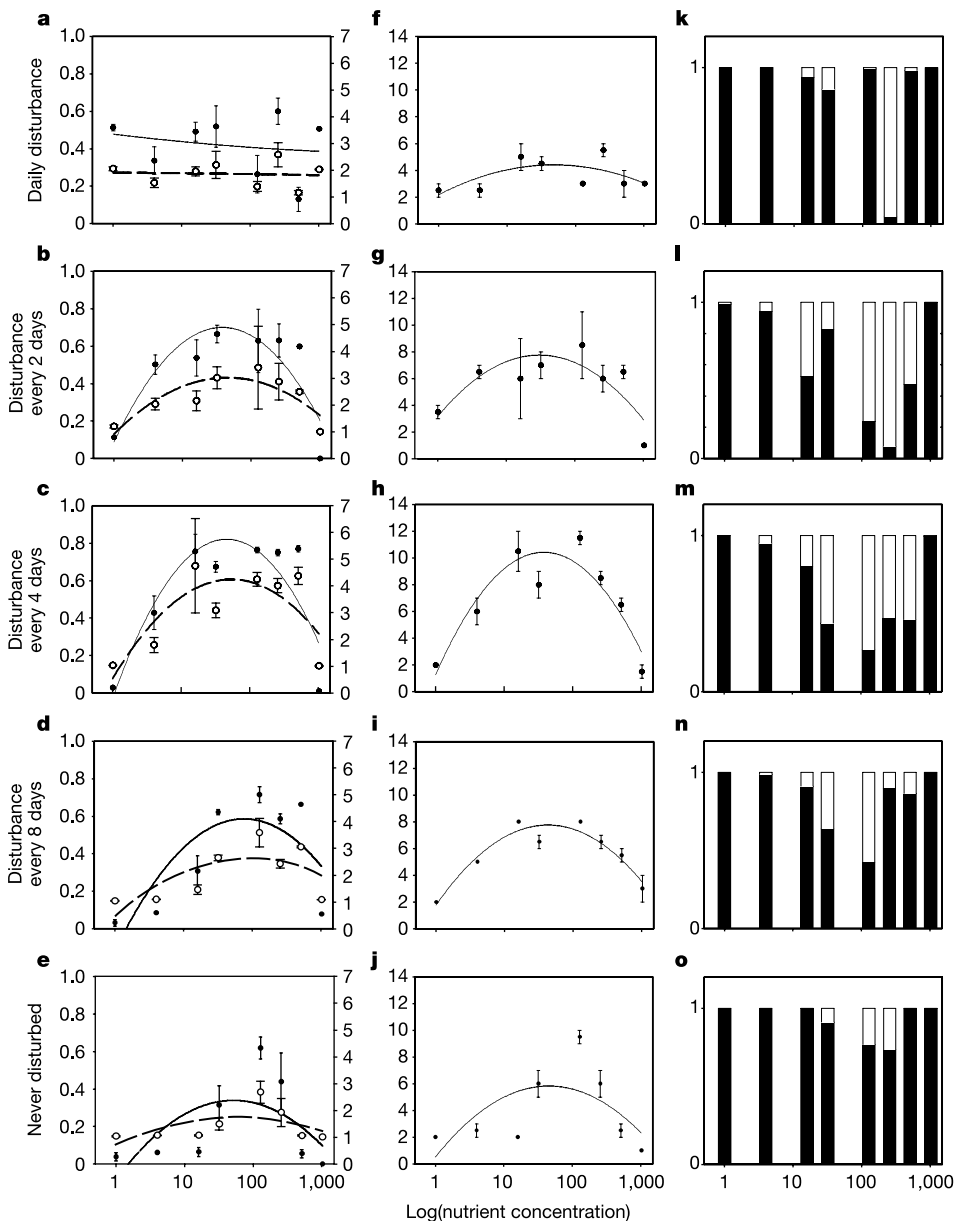


Figure 2 Diversity in relation to productivity, according to disturbance regime. **a–e**, Diversity expressed as $1 - \lambda$ (solid line) and $1/\lambda$ (dashed line) (± 1 standard error). **f–j**, Diversity expressed as richness; that is, the number of distinct colony morphs (± 1

standard error). **k–o**, The frequency of smooth morphotypes (filled sections) relative to the frequency of all other types, at each level of productivity.

Table 1 Analysis of covariance for the experiment

Source	Degrees of freedom	Mean square	F statistic	P-value	Parameter estimate ($\pm$ standard error)	Partial r^2
Productivity						
Linear	1	0.138	3.38	0.070	0.35 ± 0.19	0.06
Quadratic	1	0.533	13.1	0.001	-0.15 ± 0.04	0.26
Disturbance						
Linear	1	0.023	0.57	0.451	0.42 ± 0.55	0.07
Quadratic	1	0.164	4.03	0.048	-0.62 ± 0.31	0.07
Productivity $\times$ disturbance						
Linear	1	0.076	1.88	0.175	—	—
Quadratic	1	0.028	0.70	0.406	—	—
Residual	71	0.041	—	—	—	—

The dependent variable is diversity, measured as the complement of Simpson's index. Performing the same analysis using the reciprocal of Simpson's index gives comparable results.

competition or through the total production of individuals (carrying capacity) associated with a given patch or resources^{19,23,26,27}. The results depicted in Fig. 3 allow us to distinguish between these alternatives. If differences in competitive ability arise through changes in carrying capacities alone, then the population densities of the two types in pure culture should diverge through time, with smooth becoming more abundant than wrinkly spreader. Our results do not support this prediction (Fig. 3a–e), suggesting that the competitive superiority of smooth across a gradient of productivity stems from changes in the relative fitness of wrinkly spreader associated with the overproduction of cellulose rather than through changes to the relative carrying capacities of the two types.

Given the large population sizes (over 7×10^6 cells ml^{-1}) and asexual nature of reproduction in these populations, genetic drift and hybridization among incipient niche specialists can be eliminated as explanations for the variance in diversity observed in our experiment. An alternative explanation is that the extent of ecological opportunity, and so the strength of diversifying selection promoting diversification, decreases at the extremes of productivity and disturbance. A similar mechanism has been used to explain ecological patterns of diversity along productivity gradients in plant communities²⁸, and it may account for the inability of wrinkly spreader to invade a population of smooth morphotypes from rare at low productivity and under frequent disturbance (mean fitness $\pm$ 95% confidence limits of a standard laboratory genotype of wrinkly spreader after 1 day of competition against SBW25 at $0.00781 \times$ normal resource concentration and introduced at a ratio by volume of 1:100: 1.15 ± 1.11). However, wrinkly spreader is capable of invading from rare a smooth morphotype population under infrequent disturbance regimes¹⁸ and at high productivities in overnight competitions (mean fitness $\pm$ 95% confidence limits as above at $8 \times$ normal resource concentration: 1.86 ± 0.27), suggesting that the ecological opportunity conferred through the combination of smooth morphotype population growth and spatial structure persists when disturbance is rare and productivity high.

Ecological gradients of productivity and disturbance may thus constrain the equilibrium level of diversity achieved after adaptive radiation by decoupling evolutionary diversification from ecological coexistence, rather than through changes to the extent of ecological opportunity. The cause of this decoupling in the *P. fluorescens* radiation stems from the pleiotropic costs of niche specialization¹⁰ and manifests as costs of adaptation associated with resource competition in a spatially structured environment. This mechanism may provide a simple explanation for the variation in size between replicate adaptive radiations: large and spectacular radiations imply the existence of abundant ecological opportunity but are likely to be rare because the conditions promoting diversification may be different from those ensuring its maintenance. It is interesting to note that the pattern of species diversity in at least two well-known replicate radiations—among islands in the Hawaiian *Tetragrathia* spiders²⁹ and the Galapagos finches (see Supplementary

Fig. 2)—is unimodal, peaking on islands of intermediate age, in a manner reminiscent of our results. It is tempting to interpret these patterns along similar lines, although a variety of other factors such as dispersal³, sexual selection¹³, loss of habitable area¹⁴ or population

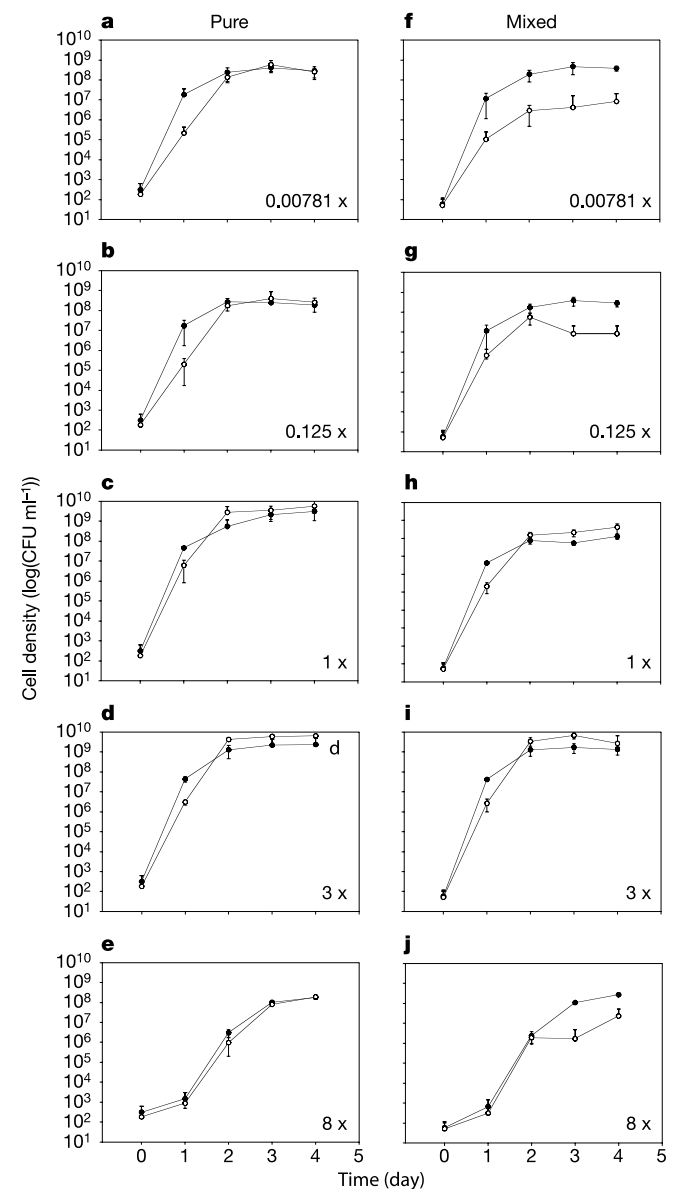


Figure 3 Population density ($\log(\text{CFU ml}^{-1})$) of smooth (filled circles) and wrinkly spreader (open circles) morphotypes over 4 days across a productivity gradient. **a–e**, Pure culture. **f–j**, Mixed culture. Error bars are 95% confidence limits.

persistence⁷ may further contribute to the variance in species diversity between replicate radiations in these more complex systems. Nevertheless our results, taken together with those of a recent study demonstrating that a further cost of adaptation may be to limit diversification itself³⁰, provides strong support for the importance of constraints in governing the size of adaptive radiations. □

Methods

Diversification across productivity and disturbance gradients

Following previous work^{17,18}, a gradient of productivity was constructed through serial dilution of a standard growth medium across a threefold range of nutrient concentrations from $8 \times$ to $0.00758 \times$ normal for a total of eight productivity levels. Global disturbances were imposed by transferring a 6- μ l aliquot (approximately 10^7 cells) from a microcosm that had been vigorously vortexed for approximately 45 s to obtain a completely mixed population into 6 ml of fresh medium daily, every 2 days, every 4 days, every 8 days, or not at all, over the course of 16 days. Cultures were plated and scored for the frequency of different colony morphologies at the end of the experiment. Diversity was estimated as the complement and reciprocal of Simpson's index, $1 - \lambda$ and $1/\lambda$, respectively, where $\lambda = \sum p_i^2$ and p_i is the frequency of each colony type counted from approximately 100 colonies and used in all statistical analyses. Final population sizes varied between 7×10^6 to 2×10^9 cells ml^{-1} , making sampling effects an unlikely explanation for our results. Note that the disturbance regime allows for approximately ten generations of growth per transfer, so frequently disturbed populations will have undergone more selection than less frequently disturbed populations. This provides a likely explanation for the high diversity at the extreme productivity levels under the daily transfer regime (Fig. 1), which was due to the presence of different kinds of smooth genotypes (Fig. 2a, f, k).

Despite obvious and dense growth in the microcosms at $8 \times$ standard nutrient concentration in the never-transferred disturbance regime, we were unable to obtain viable cells on plates from fresh cultures. Viable cells were present in the culture, however, as aliquots from frozen samples grew when inoculated into fresh medium. To estimate diversity, we allowed frozen samples of cells to recover in standard concentrations of liquid King's B medium for approximately 7 h before plating. We observed only smooth colony morphotypes from these recovered populations. Data from these cultures are treated as missing values in the statistical analysis but have been included in Figs 1 and 2 for clarity.

Population density in pure and mixed cultures

Pure and mixed cultures of the ancestral smooth (SBW25) and a derived wrinkly spreader genotype were inoculated at low density ($n \pm \text{s.e.m.}$; smooth 311 ± 58 ; wrinkly spreader 180 ± 51) into static microcosms at five productivity levels ($0.00781 \times$, $0.125 \times$, $1 \times$, $3 \times$ and $8 \times$ standard). Three replicates of each genotype and mixture were grown at 28°C and destructively sampled every day for 4 days. We measured the population size of both genotypes by counting the number of colony-forming units (CFU) at known dilution in four 10- μ l drops on standard KB agar plates.

Received 17 September; accepted 6 August 2004; doi:10.1038/nature02923.

1. Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, New York, 1953).
2. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
3. Givnish, T. J. in *Molecular Evolution and Adaptive Radiation* (eds Givnish, T. J. & Sytsma, K. J.) 1–54 (Cambridge Univ. Press, Cambridge, 1997).
4. Conway Morris, S. in *Evolution of Biological Diversity* (eds Magurran, A. E. & May, R. M.) 283–321 (Oxford Univ. Press, Oxford, 1999).
5. Meyer, A. Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends Ecol. Evol.* **8**, 267–305 (1993).
6. Williams, E. E. The origin of faunas—evolution of lizard congeners in a complex island fauna: a trail analysis. *Evol. Biol.* **6**, 47–89 (1972).
7. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
8. Huston, M. A. *Biological Diversity* (Cambridge Univ. Press, Cambridge, 1994).
9. Connell, J. H. Diversity in tropical rainforests and coral reefs. *Science* **199**, 1302–1310 (1978).
10. Maclean, R. C., Bell, G. & Rainey, P. B. The evolution of a pleiotropic fitness trade-off in *Pseudomonas fluorescens*. *Proc. Natl Acad. Sci. USA* **101**, 8072–8077 (2004).
11. Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998).
12. Day, T. & Young, K. A. Competitive and facilitative evolutionary diversification. *Bioscience* **54**, 101–109 (2004).
13. Seehausen, O., van Alphen, J. J. M. & Witt, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811 (1997).
14. Losos, J. B. & Schluter, D. Analysis of an evolutionary species-area relationship. *Nature* **408**, 847–850 (2000).
15. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421**, 259–264 (2003).
16. Travisano, M. & Rainey, P. B. Studies of adaptive radiation using microbial model systems. *Am. Nat.* **156**, S35–S44 (2000).
17. Kassen, R. *et al.* Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* **406**, 508–512 (2000).
18. Buckling, A. *et al.* Disturbance and diversity in experimental microcosms. *Nature* **408**, 961–964 (2000).
19. Riebesell, J. F. Paradox of enrichment in competitive systems. *Ecology* **55**, 183–187 (1974).
20. Kondoh, M. Unifying the relationships of species richness to productivity and disturbance. *Proc. R. Soc. Lond. B* **268**, 269–271 (2001).
21. Worm, B. *et al.* Consumer versus resource control of species diversity and ecosystem functioning. *Nature* **417**, 848–851 (2002).

22. Levene, H. Genetic equilibrium when more than one niche is available. *Am. Nat.* **87**, 331–333 (1953).
23. Maynard Smith, J. & Hoekstra, R. Polymorphism in a varied environment: how robust are the models? *Genet. Res.* **35**, 45–57 (1980).
24. Rainey, P. B. & Rainey, K. Evolution of cooperation and conflict in experimental populations of bacteria. *Nature* **425**, 72–74 (2003).
25. Spiers, A. J. *et al.* Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Mol. Microbiol.* **50**, 15–27 (2003).
26. Holt, R. D. Adaptive evolution in source-sink environments: direct and indirect effects of density-dependence on niche evolution. *Oikos* **75**, 182–192 (1996).
27. Kawecki, T. J., Barton, N. H. & Fry, J. D. Mutational collapse of fitness in marginal habitats and the evolution of ecological specialization. *J. Evol. Biol.* **10**, 407–429 (1997).
28. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, 1982).
29. Gillespie, R. Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303**, 356–359 (2004).
30. Buckling, A., Wills, M. A. & Colegrave, N. Adaptation limits diversification of experimental bacterial populations. *Science* **302**, 2107–2109 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. Bell, T. Day, M. Doebeli, B. Kerr, J. Pannell and D. Schluter for critical reviews of the manuscript and discussion. A. Gunn and J. Stansfield provided technical assistance. This work was supported by grants from NSERC (Canada) and St Hugh's College, Oxford, to R.K. and BBSRC (UK) and NERC (UK) to P.B.R.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (rkassen@uottawa.ca).

Megabase deletions of gene deserts result in viable mice

Marcelo A. Nóbrega*, Yiwen Zhu*, Ingrid Plajzer-Frick, Veena Afzal & Edward M. Rubin

DOE Joint Genome Institute Walnut Creek, California 94598, USA, and Genomics Division Lawrence Berkeley National Laboratory Berkeley, California 94720, USA

* These authors contributed equally to this work

The functional importance of the roughly 98% of mammalian genomes not corresponding to protein coding sequences remains largely undetermined¹. Here we show that some large-scale deletions of the non-coding DNA referred to as gene deserts^{2–4} can be well tolerated by an organism. We deleted two large non-coding intervals, 1,511 kilobases and 845 kilobases in length, from the mouse genome. Viable mice homozygous for the deletions were generated and were indistinguishable from wild-type littermates with regard to morphology, reproductive fitness, growth, longevity and a variety of parameters assaying general homeostasis. Further detailed analysis of the expression of multiple genes bracketing the deletions revealed only minor expression differences in homozygous deletion and wild-type mice. Together, the two deleted segments harbour 1,243 non-coding sequences conserved between humans and rodents (more than 100 base pairs, 70% identity). Some of the deleted sequences might encode for functions unidentified in our screen; nonetheless, these studies further support the existence of potentially 'disposable DNA' in the genomes of mammals.

The genome of an organism is frequently referred to as its 'book of life'⁵. It remains unclear, however, whether this is an information-dense book, in which every page is required for the proper telling of the story, or whether some of it is disposable, without an impact on the story line. For coding regions, the general necessity of maintaining most genes in animal genomes has been established in many studies. Although gene inactivation can sometimes fail to result in

detectable phenotypes, this is usually related to the removal of genes with redundancy elsewhere in the genome^{6,7}. However, our understanding of the functional importance of the non-coding sequences that represent the bulk of the genome has been only minimally examined⁸. To explore the activity of non-coding DNA on a large scale, we removed extended regions of non-coding DNA from the mouse genome, to determine its impact on the organism.

We selected two regions for deletion, a 1,817 kilobase (kb) gene desert mapping to mouse chromosome 3, and a second region, 983 kb in length, mapping to mouse chromosome 19 (Fig. 1). Orthologous gene deserts of about the same size are present on human chromosomes 1p31 and 10q23, respectively. No striking sequence signatures such as repeat content or nucleotide composition distinguish these two selected gene deserts from other regions of the genome, except for their lack of annotated genes and lack of evidence of transcription (see Supplementary Information). Together, the two selected regions contain 1,243 human–mouse conserved non-coding elements (more than 100 base pairs (bp), 70% identity), also similar to genome averages, whereas no ultra-conserved elements⁹ or sequences conserved to fish (more than 100 bp, 70% identity) are present.

We generated the deletions by coupling genomic targeting and Cre–lox technologies^{10,11} (Fig. 1b). Vectors carrying loxP sites were consecutively integrated into the mouse genome at the boundaries of each gene desert in embryonic stem cells¹². The boundaries for the deletions ranged from 51 to 255 kb from the nearest gene bracketing the desert, permitting proximate regulatory elements

nearby the flanking genes to remain intact (Fig. 1). Cre-mediated recombination in embryonic stem cells resulted in the deletion of a 1,511-kb segment on the mouse chromosome 3 desert (termed *del*^{MMU3}), and a 845-kb deletion on the mouse chromosome 19 desert (termed *del*^{MMU19}) (Fig. 1). We subsequently generated mice carrying the deletions, first as chimaeras and, after successfully transmitting the deletions through the germline, as heterozygous *del*^{MMU3}/+ and *del*^{MMU19}/+ mice. The heterozygous deletion mice, appearing phenotypically normal, were intercrossed to generate homozygous deletion mice (*del*^{MMU3}/*del*^{MMU3} and *del*^{MMU19}/*del*^{MMU19}). Because the homozygous deletion mice for both deletions were viable, we next examined a large cohort of live-born offspring from heterozygous deletion intercrosses, to determine whether the deletions had an impact on embryonic development. Genotyping 292 offspring of the *del*^{MMU3}/+ intercross and 318 offspring of the *del*^{MMU19}/+ intercross revealed an approximate 1.0:2.0:1.0 ratio of wild-type to heterozygous to mutant homozygous mice obtained for both the MMU3 and MMU19 deletions, indicating that the deletions do not result in embryonic lethality (see Supplementary Information).

We next explored the overall fitness and a variety of phenotypic parameters measured in the homozygous deletion mice, compared with controls. Monitoring post-natal survival rates for 25 weeks for each of the deletions revealed no differences between *del*^{MMU3}/*del*^{MMU3} or *del*^{MMU19}/*del*^{MMU19} and their respective wild-type littermates (Fig. 2a). Neither genomic deletion elicited measurable growth retardation either, as shown by similar growth curves

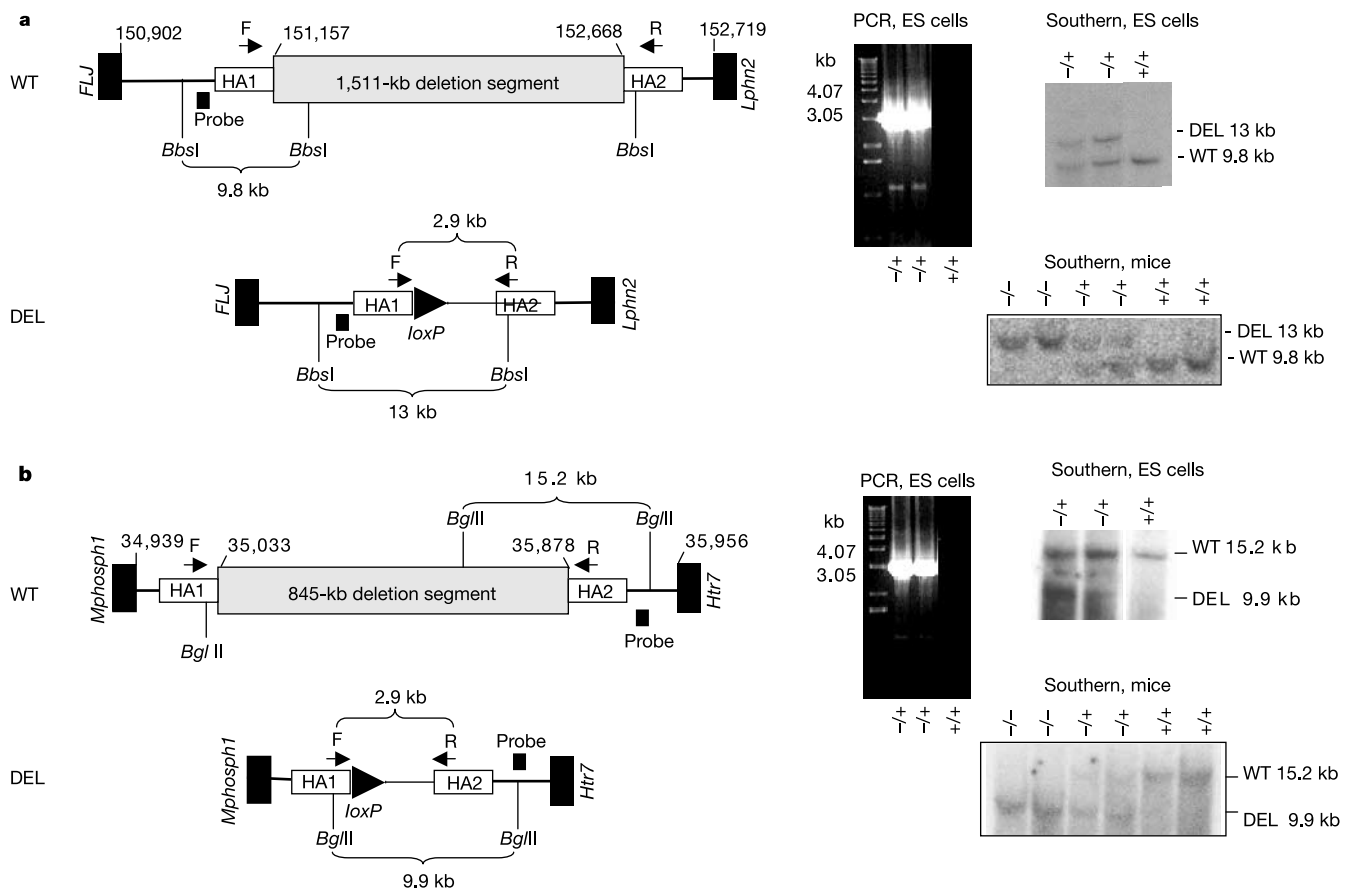


Figure 1 Design of genomic deletions in gene deserts MMU3 and MMU19, displaying a wild-type (WT) and a deletion (DEL) allele for each desert. **a**, MMU3 desert; **b**, MMU19 desert. Left: both deserts (figures not drawn to scale) and the genes bracketing each deletion are illustrated with coordinates (in kb) from the NCBI mouse build 32. The points of integration of target vector homology arms (HA1 and HA2) are shown. Primers (F and R)

used to screen embryonic stem (ES) cells by PCR for the presence of the deletions are indicated by arrows flanking each deleted segment. Right, probes used in the Southern hybridization for identification of embryonic stem cells and mice carrying the deletions are shown, together with the restriction sites and the sizes of WT and DEL alleles for each desert. PCR and Southern hybridization of each allele are illustrated.

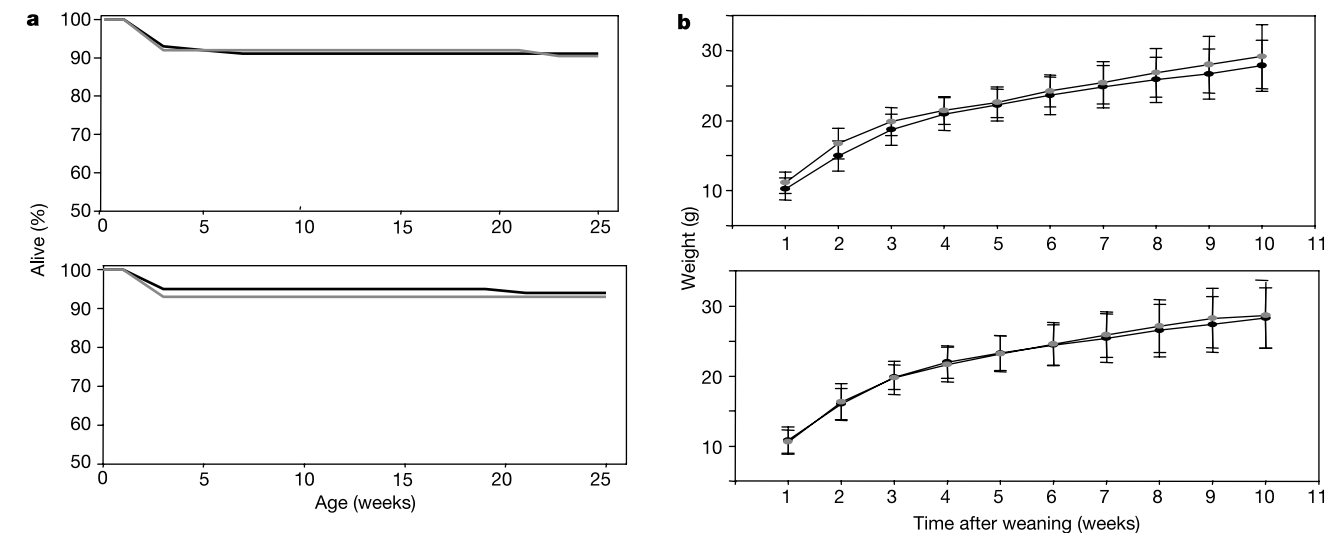


Figure 2 Survival and growth curves of mice carrying the MMU3 and MMU19 deletions. **a**, Survival curves (upper, MMU19, $n = 155$; lower, MMU3, $n = 137$); **b**, growth curves (upper, MMU19, $n = 47$; lower, MMU3, $n = 53$). In **a**, grey lines represent (+/+) and

black lines (—/—) in **b**, grey dots represent (+/+) and black dots (—/—). Error bars indicate s.d.

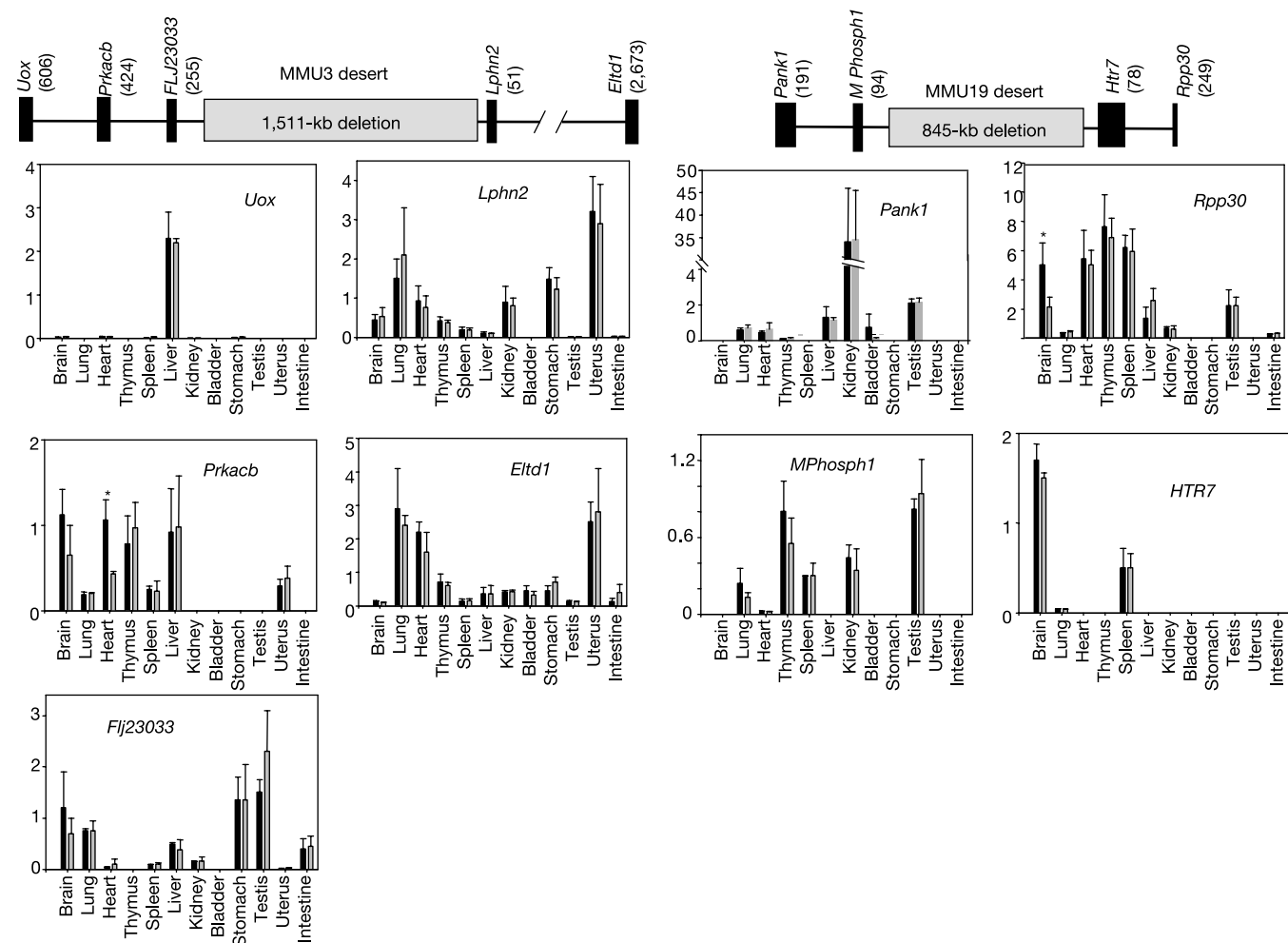


Figure 3 Expression of genes bracketing the chromosomal deletions in a panel of 12 tissues. Distances of each gene from the closest end of the deletion are given in parentheses (in kb). Vertical axes correspond to the ratio of the expression of the corresponding gene to 18S RNA expression at each given tissue. Asterisks denote

statistically significant differences between homozygous deletion mice and wild-type littermates ($P < 0.05$). Black columns, (+/+); grey columns, (-/-). Error bars indicate s.d.

obtained for each of the four groups, del^{MMU3}/del^{MMU3} , del^{MMU19}/del^{MMU19} and their wild-type littermates (Fig. 2b). In addition, we measured 18 different clinical chemistry tests representing both general and specific plasma parameters in the same groups of mice, also with no significant differences detected between del^{MMU3}/del^{MMU3} or del^{MMU19}/del^{MMU19} and their wild-type littermate controls (see Supplementary Information). We then performed visual and pathological examinations on multiple organs from the various groups of mice at 6 months of age, including brain, thymus, heart, lungs, liver, spleen, stomach, intestine, kidney, urinary bladder, uterus, ovaries and testicles. No morphological abnormalities, evidence of abnormal growth or tissue degeneration were observed in del^{MMU3}/del^{MMU3} or del^{MMU19}/del^{MMU19} mice. Organ mass was similar in both groups of deletion mice and their wild-type littermates (see Supplementary Information).

Finding no differences in any of the whole organism phenotypes tested, we subsequently explored the impact of the deletions at a molecular level. The expression levels of multiple genes flanking the boundaries of the two deletions were determined in 12-week-old mice. We assayed four genes bracketing the MMU19 deletion and five genes bracketing the MMU3 deletion by real-time quantitative polymerase chain reaction (PCR), in a panel of 12 tissues representing the overall expression patterns of each assayed gene. The tissue specificity of expression for all the genes tested was similar in homozygous deletion mice and their wild-type littermates (Fig. 3). Of the 108 quantitative expression assays (12 tissues for 9 genes), only 2 revealed detectable alterations in levels of expression. The expression of *Prkab* was reduced in the heart of del^{MMU3}/del^{MMU3} mice and *Rpp30* was reduced in the brain of del^{MMU19}/del^{MMU19} mice, compared with wild-type littermates (Fig. 3).

Given the small number of expression changes that we did observe, coupled with the limited set of tissues used for the expression assay, we performed a series of reporter assays in

transgenic mice to explore further the possible existence of regulatory sequences in the deleted segments^{13–15}. The elements selected for testing in this gain-of-function assay correspond to those with the highest degree of sequence conservation extending over the largest intervals between humans and mice¹⁶. We also sought those elements with conservation extending over the longest evolutionary timescale. Although non-coding conservation in the MMU19 desert was shared only between mammals, the MMU3 desert contained human–mouse conserved elements also shared with chicken (75 elements) and frog (5 elements). From the MMU19 desert we picked five human–mouse conserved elements representing the most conserved sequences between these species (more than 180 bp, 90% identity) for the *in vivo* assay. The ten elements chosen from desert MMU3 (more than 400 bp, 90% identity) included all five sequences that are conserved across humans, rodents, chicken and frog, and five that are conserved across humans, rodents and chicken only (Fig. 4).

We cloned each element in a reporter vector¹⁷, injected it into fertilized mouse oocytes and assayed for the presence of β -galactosidase activity in the resulting embryos at 14.5 days *post coitum* (E14.5); 8–16 independent transgenic mice that were generated for each of the 15 elements were examined. Of the elements tested, only one, located within the desert MMU3, reproducibly drove β -galactosidase expression in a set of tissues that included mammary glands and abdominal muscles (Fig. 4). It is noteworthy that this element is conserved deep in the vertebrate lineage, including human, mouse, chicken and frog. The small fraction of elements conserved across several vertebrates but not fish with enhancer activity in this study (1 of 15) contrasts with the results obtained when human–fish conserved non-coding sequences were previously tested with the same *in vivo* assay^{18,19}. In those studies a significant fraction of human–fish conserved non-coding sequences present in gene deserts were shown to be functional, with 5 of 7 elements in

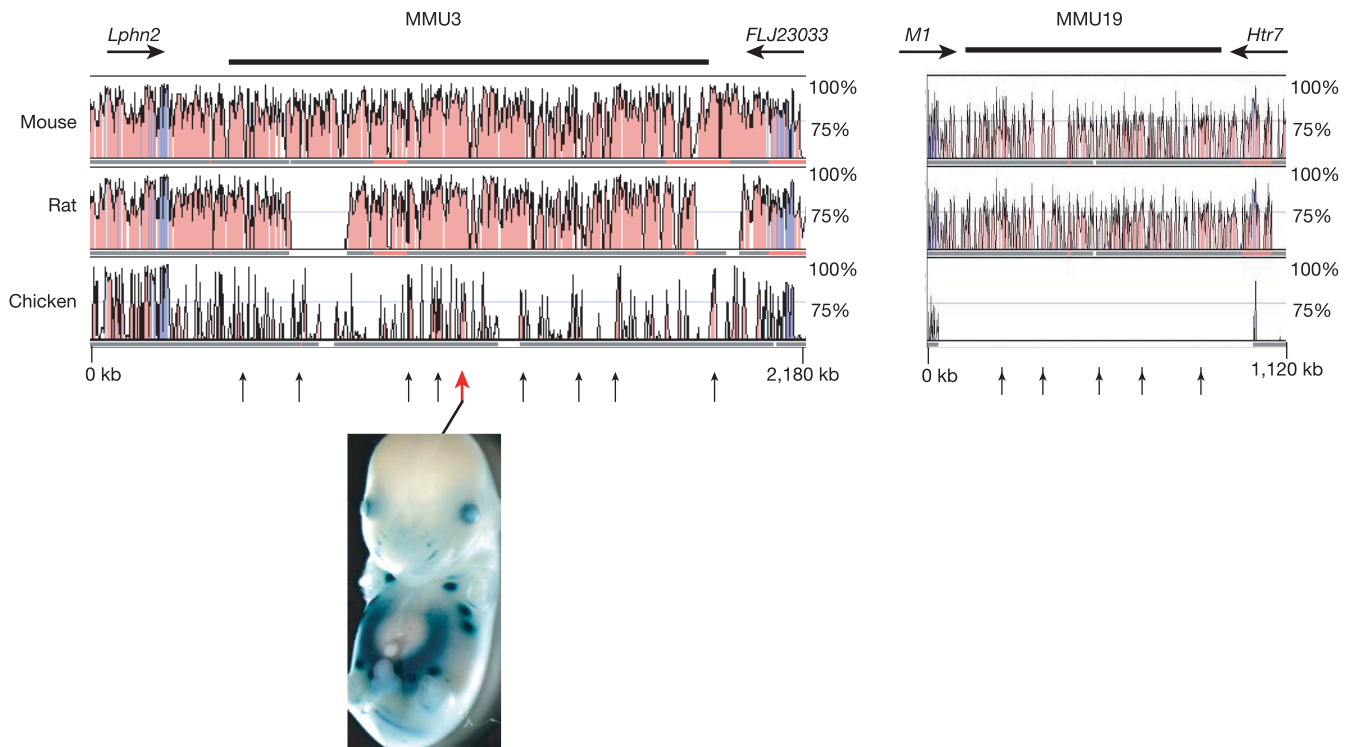


Figure 4 Conservation profile between human, mouse, rat and chicken. Alignments were obtained with Vista (<http://www-gsd.lbl.gov/VISTA>), with the human sequence as reference in the horizontal axis. Exons of genes bracketing the deserts are denoted in blue, whereas non-coding sequence conservation is shown in red. A bar over each plot

illustrates the deleted segments. Sequence elements selected for the *in vivo* mouse transgenic assay are shown by arrows at the bottom of each conservation plot. The element containing enhancer activity at E14.5 is highlighted by a red arrow, with a representative embryo showing the expression pattern of the element displayed.

one study¹³ and 22 of 29 in a second study (E.M.R., unpublished data) demonstrating enhancer activity.

The deletions made in this study are the largest reported viable homozygous deletions in mice and support the existence of potentially 'disposable DNA' in mammalian genomes. In assessing the impact of these deletions on the engineered mice, it is important to acknowledge that our ability to phenotype an organism will always miss some features, no matter how detailed the analysis. It is possible—even likely—that the animals carrying the megabase-long genomic deletions do harbour abnormalities undetected in our assays, which might affect their fitness in some other timescale or setting than those assayed in this study. In addition, just as gene redundancy can mask the impact of gene inactivation in animal models, regulatory sequence redundancy²⁰ existing outside the deleted intervals might mask the impact of the removal of such elements by the deletions. Nonetheless, the lack of detectable phenotypes in these mice raises the possibility that the mammalian genome is not densely encoded and that significant reductions in genome size may be tolerated. This concept has recently been supported by the demonstration that, in 'normal' humans, large-scale genomic deletions are a common source of polymorphism, generally occurring in gene-poor regions, some of which contain sequence conservation across mammal species^{21,22}. Linked to this, the extensive degree of non-coding conservation in the deleted intervals in our mouse study and also in the human studies brings into question the functionality, if any, of many of the large number of non-coding sequences shared between mammals. □

Methods

Genomic coordinates of the gene deserts

All coordinates were obtained with the NCBI mouse build 32 (<http://genome.ucsc.edu>) as reference. A more detailed annotation of each segment is given in the Supplementary Information. The MM3 deleted segment was chr3:151,157,238–152,668,920; the MMU19 deleted segment was chr19:35,033,162–35,878,431.

Generation of the targeting vectors

Two targeting vectors were generated, in house, for each deletion. One targeting vector consists of a *loxP* site coupled to a neomycin-resistance gene and HSV-tk (*ploxP-Neo-tk-2-BRI*) and the other vector consists of a *loxP* site coupled to a hygromycin-resistance gene and HSV-tk (*ploxP-Hyg-tk-BH*). Homologous arms, 5,700 to 7,000 bp in length, were generated by amplification from 129Sv mouse genomic DNA, with the PCR products cloned initially into a pCR2.1 vector (TOPO XL PCR cloning kit; Invitrogen), and subsequently cloned into each corresponding targeting vector.

Creation of the genomic deletions *in vitro*

Constructs were linearized and electroporated (20 µg) into mouse ESJ embryonic stem cells (129Sv genetic background). Cells were electroporated with either *ploxP-Neo-tk-2-BRI* or *ploxP-Hyg-tk-BH* constructs and selected under G418 (180 µg ml⁻¹) or hygromycin B (150 µg ml⁻¹), respectively, for 8 days. Resistant colonies were picked and expanded on 96-well plates and screened by both PCR and Southern blot hybridization, with probes targeting the vicinity of the homologous arms. Clones that were correctly double-targeted were electroporated with 20 µg of the Cre recombinase-expressing plasmid pBS185 (ref. 23). At this stage one cannot tell whether the vectors are integrated in one chromosome (*cis*) or in both homologous chromosomes (*trans*). Only in cells with both vectors integrated *in cis* will Cre-mediated genomic deletion with a loss of both copies of HSV-tk occur. Embryonic stem cells that had undergone recombination and deletion were identified by selection for hygromycin sensitivity and 1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-iodouracil resistance, which selects for colonies in which the deletion removed the HSV-tk genes. Resistant colonies were screened both by PCR, using primers outside the deleted region, and by Southern blot hybridization to identify those carrying the designed genomic deletion (Fig. 1). All primers used for PCR and Southern blot hybridization are given in the Supplementary Information.

Generation of mice carrying the genomic deletions

The correctly targeted mouse embryonic stem cells were injected into C57BL/6 blastocysts, and mice were generated as described previously²⁴. Male chimaeric mice were bred with C57BL/6 females for germline transmission of the deletions and for the generation of heterozygous deletion mice. Heterozygous breeding generated the homozygous deletion mice. Although introgressing the deletions into multiple isogenic backgrounds might reveal phenotypic differences specific to certain backgrounds, we decided to use a more robust hybrid background, minimizing potential genetic background confounding factors. Thus, no subsequent backcrosses were performed, and all mice reported in this study carried a hybrid of C57BL/6 and 129/SvEv backgrounds.

Genotyping

Genomic DNA was extracted from a 1-cm section of tail that was incubated overnight in lysis buffer (containing 100 mM Tris-HCl pH 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl and 50 µg Proteinase K) at 55 °C. Genotyping was performed by PCR, with primers outside the deleted segments and within the targeting vector (see Supplementary Information for details).

Clinical chemistry

After an overnight fast, about 200–300 µl of blood was collected from the tail vein of 6-month-old mice. Serum parameters were determined by following a protocol used by the NHLBI PhysGen PGA as described (<http://pga.mcw.edu/>).

Real-time quantitative PCR

Total RNA was extracted, with TRIzol reagent (Invitrogen), from each of the 12 organs tested, and pooled (four males and four females). After the treatment of total RNA with Promega RQ1 RNase-Free DNase, reverse transcription was performed with SuperScript First-Strand Synthesis kit (Invitrogen). Quantitative real-time PCR was performed with SYBR Green PCR Master Mix (Applied Biosystems). An 18S RNA standard internal control (Ambion) was used. A 4:6 ratio of 18S primer pair to 18S competitors was adopted. All procedures and calculations were performed in accordance with manufacturers' recommendations.

Mouse transgenic reporter assay

Adopting a conservation criterion of at least 100 bp and 70% identity, 1,243 human–mouse conserved elements were identified in the deleted segments. At this level of stringency, it is likely that most of these conserved sequences fall into the roughly 5% of the genome that is thought to be evolving under selective pressure^{25,26}. From these 1,243 elements we identified 15 that displayed the strongest conservation parameters for testing *in vivo*. Each conserved non-coding sequence was amplified from human DNA (Invitrogen) by PCR, with *Pst*I, *Hind*III or *Xho*I linkers on each end, and cloned into *phsp68/LacZ*. The constructs were purified and injected into pronuclei as described previously²⁷. Embryos were harvested at E14.5, and transgenic embryos were identified by PCR of β-galactosidase, using DNA from yolk sac, with the following primers: F, 5'-TTTCCATGTTGCCACTCGC-3'; R, 5'-AACGGCTTGCCGTTTCAGCA-3'. Expression of β-galactosidase was assayed in all embryos, using 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal; Sigma), as described previously²⁸.

Received 2 June; accepted 8 September 2004; doi:10.1038/nature03022.

- Bentley, D. R. Genomes for medicine. *Nature* **429**, 440–445 (2004).
- Dunham, A. *et al.* The DNA sequence and analysis of human chromosome 13. *Nature* **428**, 522–528 (2004).
- Scherer, S. W. *et al.* Human chromosome 7: DNA sequence and biology. *Science* **300**, 767–772 (2003).
- Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Jasny, B. R. & Kennedy, D. The human genome. *Science* **291**, 1153 (2001).
- Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
- Winzler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Nelson, C. E., Hersh, B. M. & Carroll, S. B. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol.* **5**, R25.1–R25.15 (2004).
- Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
- Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
- Yu, Y. & Bradley, A. Engineering chromosomal rearrangements in mice. *Nature Rev. Genet.* **2**, 780–790 (2001).
- Zhu, Y. *et al.* Genomic interval engineering of mice identifies a novel modulator of triglyceride production. *Proc. Natl Acad. Sci. USA* **97**, 1137–1142 (2000).
- Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
- Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
- Kimura-Yoshida, C. *et al.* Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. *Development* **131**, 57–71 (2004).
- Ovcharenko, I., Loots, G. G. & Stubbs, L. Interpreting mammalian evolution using Fugu genome comparisons. *Genomics* (in the press).
- Kothary, R. *et al.* Inducible expression of an hsp68–lacZ hybrid gene in transgenic mice. *Development* **105**, 707–714 (1989).
- Boffelli, D., Nobrega, M. A. & Rubin, E. M. Comparative genomics at the vertebrate extremes. *Nature Rev. Genet.* **5**, 456–465 (2004).
- Zerucha, T. *et al.* A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J. Neurosci.* **20**, 709–721 (2000).
- Frazier, K. A. *et al.* Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Res.* **14**, 367–372 (2004).
- Iafate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
- Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- Sauer, B. & Henderson, N. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol.* **2**, 441–449 (1990).
- Hogan, B. C. B. & Lacy, E. *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Plainview, New York, 1994).
- Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).

26. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
27. Anderson, J. P. *et al.* HRC is a direct transcriptional target of MEF2 during cardiac, skeletal, and arterial smooth muscle development *in vivo*. *Mol. Cell. Biol.* **24**, 3757–3768 (2004).
28. Dodou, E., Xu, S. M. & Black, B. L. *me2c* is activated directly by myogenic basic helix–loop–helix proteins during skeletal muscle development *in vivo*. *Mech. Dev.* **120**, 1021–1032 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank I. Ovcharenko, G. Loots and J. Schwartz for help with the identification and annotation of the gene deserts; D. Boffelli, L. Pennacchio, N. Ahituv, J. Bristow and other Rubin laboratory members for suggestions and criticisms on the manuscript; and H. Jacob for providing the clinical chemistry assays. Research was conducted at the E. O. Lawrence Berkeley National Laboratory and at the Joint Genome Institute, with support by grants from the Programs for Genomic Application, the NHLBI and the DOE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.M.R. (emrubin@lbl.gov).

Neural correlates of mental rehearsal in dorsal premotor cortex

Paul Cisek & John F. Kalaska

Département de physiologie, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, Québec, H3C 3J7, Canada

Behavioural and imaging studies suggest that when humans mentally rehearse a familiar action they execute some of the same neural operations used during overt motor performance^{1–9}. Similarly, neural activation is present during action observation in many of the same brain regions normally used for performance, including premotor cortex^{6–9}. Here we present behavioural evidence that monkeys also engage in mental rehearsal during the observation of sensory events associated with a well-learned motor task. Furthermore, most task-related neurons in dorsal premotor cortex exhibit the same activity patterns

during observation as during performance, even during an instructed-delay period before any actual observed motion. This activity might be a single-neuron correlate of covert mental rehearsal.

The ‘direct matching hypothesis’^{10–13} suggests that understanding the observed actions of others not only involves a cognitive analysis of sensory inputs but also evokes a representation of the action in some of the same neural circuits that are engaged when the observer performs those actions, a form of covert simulation or mental rehearsal of the observed acts^{6,7,9,14}. At the single-neuron level, support for the hypothesis is provided by a select population of ‘mirror neurons’ in the ventral premotor cortex. These neurons discharge both while a monkey performs natural actions such as reaching, grasping or biting, and when the monkey observes someone else performing those same familiar actions directed at target objects^{10–13}. These findings indicate that premotor cortex (PM) might be involved not only in overt motor performance, but also in non-motor functions such as recognizing the nature and consequences of the actions of others, and attribution of agency^{1,12,14}.

However, activation of motor circuits during mental rehearsal occurs not only during the direct observation of an ongoing action but also whenever subjects are given information specifying a particular action^{9,14,15}. This predicts that the neural mechanisms involved in selecting and planning movements should also be activated, in advance of an observed action, as soon as the observer is given information that allows them to anticipate what will happen. Evidence that mental rehearsal can include anticipatory processes is indirectly supported by the finding that the oculomotor behaviour of humans is similar when they observe someone performing a block-stacking task and when they perform the task themselves¹⁶. In particular, the direction of gaze predicts what should happen next, indicating that the observers are mentally simulating the task to predict future events instead of just following the viewed events with their eyes as they occur. However, mirror neurons do not discharge in anticipation of a potential action when either the actor or target, or both, are viewed at rest, or during the observation of pantomimed actions¹². Furthermore, mirror neurons respond only to natural actions that are directly viewed, and not to their video images¹⁷. These findings indicate a specific role for mirror neurons in the interpretation of ongoing familiar actions of a

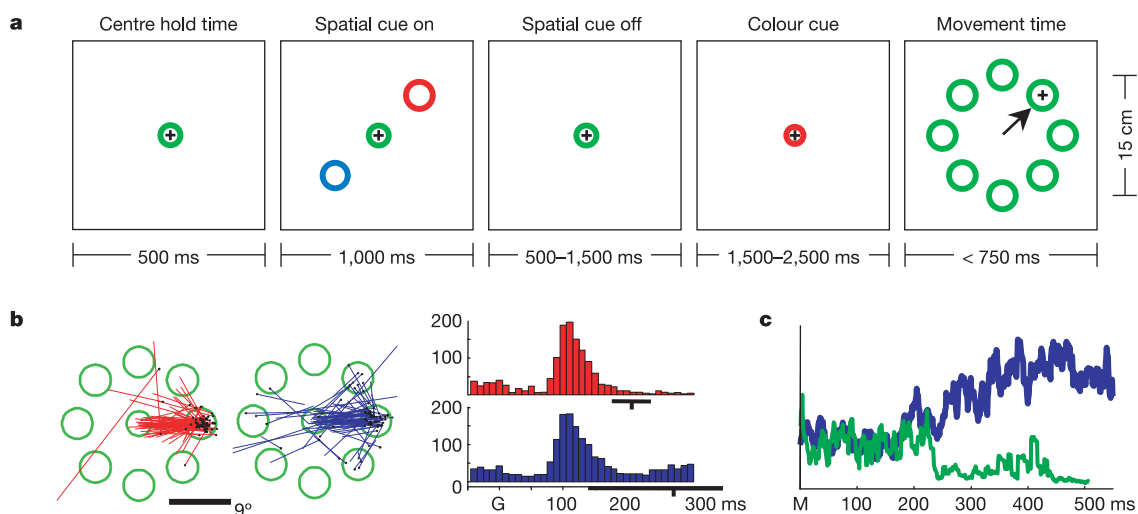


Figure 1 Behavioural data. **a**, Stimulus events in the two-target task. The cross indicates cursor location. **b**, Saccadic behaviour. Left: representative saccades 50–200 ms after GO when the right target was selected are shown as coloured lines (red, performance; blue, observation) terminating with black dots indicating fixation. Right: histograms of saccade onset relative to GO (G), during performance (red, upper) and observation (blue,

lower). Black bars indicate mean and s.d. of cursor motion onset. **c**, Electromyogram activity of mentalis (a lip muscle) during observation trials in which the experimenter moved to the correct target (blue) and to the opposite target (green). Activity is aligned on cursor motion onset (M) and averaged across trials.

directly observed actor, but they do not seem to be implicated in the prediction of impending actions or events based on arbitrary information.

In contrast, we report here that neurons in the dorsal premotor cortex (PMd) of monkeys might contribute to mental rehearsal of processes that normally occur before the onset of movement. Unlike mirror neurons, their activities predict the directionality of impending events before the onset of movement, they respond to abstract visual cues that have become associated with action only through training, and they do not require the actor to be in view. These PMd neurons emit very similar responses when the monkeys actively perform the motor task, indicating that passive observation of the artificial stimulus events might evoke a covert rehearsal of the same premotor neuronal operations that would normally lead to overt motor output.

Two monkeys learned a reaching task guided by visual stimuli presented on a computer monitor¹⁸ (Fig. 1a). During the first

instructed-delay period (spatial cue, SC), two colour-coded potential targets for movement were displayed briefly, and during the second delay period (colour cue, CC), one of them was designated as the correct target. The monkeys learned to make reaching movements in the appropriate direction in a horizontal plane, out of sight and out of the vertical plane of the screen, to move an on-screen cursor into the designated target (performance task). The monkeys usually made a saccade to the correct target just before reaching (Fig. 1b, left).

After lengthy training and weeks of neuronal recording with the performance task, we presented the monkeys with a novel situation. In a separate block of trials (observation task), the monkeys sat quietly and watched the usual task-related events unfold on the monitor while the task was performed by an unseen party. To encourage the monkeys to remain attentive to the observed events, they received a juice reward whenever the unseen party performed a trial correctly. The monkeys spontaneously exhibited two predictive

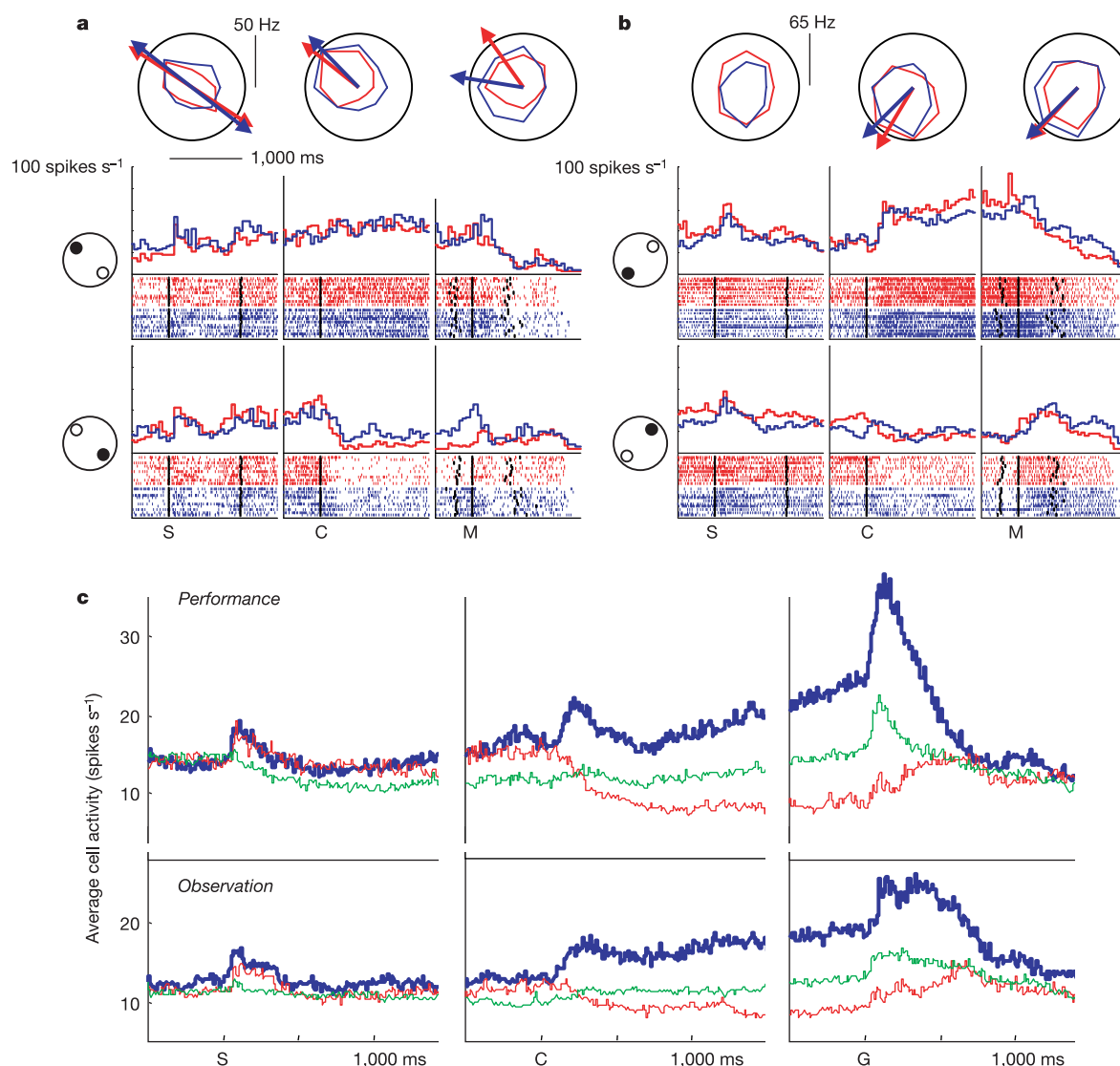


Figure 2 Neural data. **a**, Activity of a PMd cell during performance (red) and observation (blue). In the rasters, coloured marks indicate action potentials and black marks indicate behavioural events (from left to right: SC onset (S) and offset, CC onset (C), GO, movement onset (M) and offset). Data are shown for trials in which the cell's preferred target (upper) or the opposite target (lower) were specified by the CC, as indicated by the circular diagrams on the left. Polar plots show the cell's directional tuning function during SC, CC, and movement time (MT). Arrows indicate the preferred direction (or axis) of statistically

significant unimodal (or bimodal) tuning (bootstrap, $P < 0.01$). **b**, Activity of a second cell in the same format. **c**, Population histograms collected as the average activity of cells during trials to each cell's best tested direction (blue), opposite direction (red) or orthogonal directions (green). The best tested direction was defined using the preferred direction from the reaction time of performance, and only cells with significant tuning in that period were included ($N = 28$).

behaviours¹⁶, showing that they attended to the task events and knew when it was being performed correctly. First, after the GO signal but before the onset of observed cursor motion, the monkeys made a saccade to the correct target, demonstrating that their eye movements were guided by knowledge of what should happen next¹⁶. Saccade latencies were similar to those made when the monkeys performed the task themselves (Fig. 1b). Second, whenever the unseen party caused the cursor to move in the correct direction, the monkeys vigorously licked the reward tube long before the reward was delivered, but quickly stopped licking when it moved in the wrong direction (Fig. 1c), again long before the reward would have been delivered. This predictive behaviour by both monkeys demonstrated that they attended to task events, assessed their correctness¹⁵, judged the likelihood of impending reward, and acted on that judgment by licking the juice tube or not¹⁹.

The direct-matching hypothesis suggests that these behavioural phenomena are evidence of covert mental rehearsal of the learned motor task associated with the visual events, and predicts that individual neurons in arm-movement-related motor regions express qualitatively similar activity in both the observation and performance conditions. Furthermore, neural responses during observation should occur even before the onset of cursor motion and should signal the required direction of motion.

In accord with these predictions, the large majority of task-related PMd neurons recorded from both monkeys exhibited strikingly similar activity patterns during both performance and observation (Fig. 2), even before the onset of cursor motion. The responses were found in the first cell tested during observation in each animal. Of the 42 PMd cells eventually tested during observation, 38 were directionally tuned during one or more instructed-delay periods in the performance task, and 32 of 38 (84%) were also tuned during one or more instructed-delay periods in the observation task. For example, the cell in Fig. 2a discharged as soon as the target in the upper left was identified as a potential movement target during the SC period, regardless of whether the monkey intended to perform the associated actions (red rasters and histograms) or simply observed the instructional stimuli while the unseen party performed

the task (blue rasters and histograms). The cell in Fig. 2b discharged only after information was presented during the CC period that unambiguously selected the bottom left target, during both performance and observation conditions.

When the activity of PMd cells was pooled, the population signals during both performance and observation were strikingly similar (Fig. 2c). After the onset of the two SC stimuli, the activity of two populations of cells tuned to those directions (Fig. 2c, blue and red) increased relative to their activity for orthogonal directions (green). After the central colour change, activity tuned to the selected target (blue) increased while activity tuned to the unselected target (red) decreased. This showed that many PMd cells performed similar neuronal operations in response to the information provided by instructional cues during observation and performance.

Various control experiments and analyses were performed to determine whether these PMd responses simply reflected sensory driving, gaze-related or attention-related modulation, or overt motor outputs (see Supplementary Information). To summarize, it is unlikely that any of these alternative explanations could be the principal cause of the directionally tuned activity during observation.

After the GO signal, PMd cells often generated further task-related responses during observation (Fig. 2c). We investigated whether activity during the reaction time (RT), movement time (MT) and target hold time (THT) was related to the instructed cursor motion or to the motion that was actually observed. For a subset ($N = 20$) of the cells, the unseen party deliberately made errors after the GO signal, such as delaying movement or moving in the direction opposite to the correct target. Figure 3 illustrates the activity of one PMd cell, which became tuned just before and during movement. That cell showed remarkably similar responses whether the monkey (red) or the unseen party (blue) promptly performed correct responses. When the selected target was to the left, the cell became active shortly after the GO signal but before movement onset, even if the movement was delayed (Fig. 3, left, purple) and remained active until the correct leftward movement was completed. When the selected target was to the right, the cell was initially suppressed before movement onset, and then discharged only after

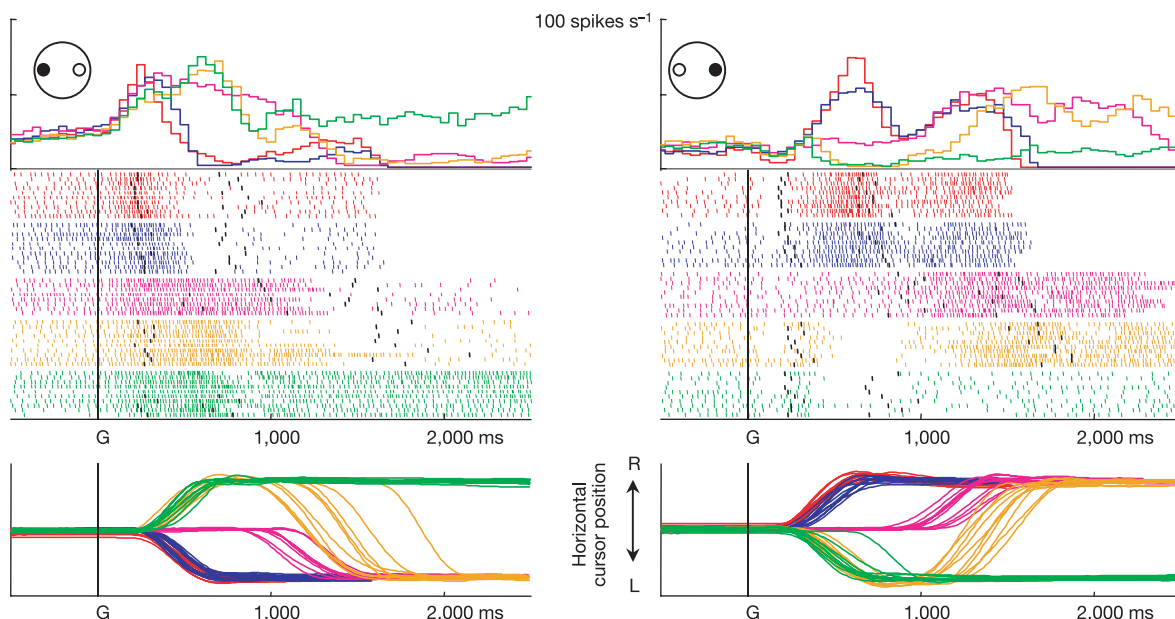


Figure 3 Activity of a PMd neuron during performance (red) and observation in trials in which the experimenter either moved promptly to the correct target (blue), delayed and then moved to the correct target (purple), moved to the opposite target and then reversed

(orange), or moved to the opposite target (green). Left: rasters, histograms and horizontal cursor position, aligned on the GO signal (G), during trials in which the correct target was on the left. Right: the same, for trials in which the correct target was on the right.

the correct rightward movement began (Fig. 3, right). In all cases, the activity of this cell during the unseen party's behavioural reaction time and for a brief time after cursor movement onset reflected the expected movement direction. Thus, even these early activity changes after the GO signal were centrally generated and were not a passive sensory response to cursor motion, which had not yet occurred. Later in the trial, cell activity often began to reflect the observed direction of cursor motion (see Supplementary Information for more quantitative analyses).

These PMd responses were very common. Our sole criterion for studying a cell during observation was that it was directionally tuned in the performance task. Nearly all (41 of 42, 98%) of the tested cells showed directional tuning during one or more of the movement periods in the observation task, and 32 of 42 (76%) showed directional tuning during one or more of the instructed-delay periods.

Although the phenomenon reported here in dorsal PM differs in several important ways from mirror neuron activity in ventral PM^{10–13,20}, it is also possible that they are functionally related. Both findings show that observation of external events can engage nominally 'motor' circuits, to generate motor representations of the actions associated with those events. This covert simulation or mental rehearsal of motor acts may contribute to nominally 'cognitive' functions underlying the assessment and understanding of observed events. For ventral premotor mirror neurons those events are natural and familiar, whereas for the PMd neurons they include sensory events whose relation to motor behaviour is acquired through learned stimulus–response associations.

Behavioural and functional imaging evidence suggests that mentally rehearsed motor actions obey the same constraints as overt actions^{3,5}, are similarly disrupted by lesions⁴, engage in a somatotopic manner⁶ many of the same brain areas (including PM) that are activated during overt motor behaviour^{2,7,9}, and facilitate the motor system in similar ways¹. The present findings indicate that mental rehearsal and overt motor planning involve very similar single-neuron operations. Mental rehearsal is still poorly understood, and may involve many of the processes associated with overt performance, ranging from representations of higher-level task goals and motor outputs to specific details of the limb movements. This study cannot distinguish the level of abstraction at which the PMd activity represented task-related events. However, when visual feedback is dissociated from the actual limb movements that produce it, PMd activity during the delay period covaries more closely with the directionality of visual feedback than with the direction of physical motor output^{21,22}. We propose that the activity reported here, in particular during the instructed-delay period, represents action in a coordinate frame related to desired cursor motions rather than specific limb movements^{21,22}. This is not a passive sensory representation. Through training, the cursor became a surrogate for the monkeys' arm, whose motion they control through arm movements. PMd generated a representation of intended motions of the controlled object (in this task, the cursor) in a coordinate framework related to the sensory feedback about the action^{21,22}, a variable of critical importance to the monkeys during both performance and observation. □

Methods

During the performance task, the monkeys controlled the motion of an on-screen cursor on a computer monitor positioned in front of them by moving the lower end of a two-degree-of-freedom pendulum in the horizontal plane with the arm contralateral to the cortical recording site. The monkeys could not see their arm, and task performance was guided entirely by the instructional cues and cursor motions on the monitor. During the observation task, the monkeys sat in the same position relative to the screen but both arms were lightly restrained, and the pendulum was moved by the experimenter who grasped it out of the monkey's sight. During the observation task, the monkeys received the usual juice reward each time the experimenter successfully completed a trial. The observation task did not form part of the training regimen before neural recordings. In both conditions, the monkey's gaze was unconstrained and eye movements were measured at 100 Hz with an infrared oculometer (Dr Bouis Devices, Karlsruhe).

During both tasks (Fig. 1a), each trial began when the on-screen cursor was moved into a central green circle (1.5 cm radius) for a 500-ms centre-hold time. Two coloured cue circles (radius 2 cm), one red and one blue, then appeared at two out of eight possible target locations on the circumference of a circle of radius 8 cm, for a spatial-cue-on period (1,000 ms) and then disappeared for a spatial-cue-off period (500–1,500 ms). Next, the central circle changed colour to either red or blue, for a colour-cue period (1,500–2,500 ms). This colour cue instructed which of the two memorized colour-coded spatial-cue locations was the selected target for the forthcoming movement. Finally, the central circle disappeared and green circles (radius 2 cm) appeared at all eight peripheral locations. This GO signal instructed the subject to move the cursor to the selected target.

Neurons were studied in both the performance and observation tasks in PMd in three hemispheres of the two monkeys. All recorded neurons were collected from the cortex along the rostral bank of the arcuate sulcus and rostral to the arcuate spur, spanning a region from 5 mm caudal to 3 mm rostral to the genu of the arcuate²³. All training, surgical and recording procedures were performed in accordance with institutional, Canadian and NIH guidelines. Further methodological details can be found in the Supplementary Information.

Received 2 August; accepted 13 September 2004; doi:10.1038/nature03005.

- Jeannerod, M. The 25th Bartlett Lecture. To act or not to act: perspectives on the representation of actions. *Q. J. Exp. Psychol. A* **52**, 1–29 (1999).
- Jeannerod, M. & Decety, J. Mental motor imagery: a window into the representational stages of action. *Curr. Opin. Neurobiol.* **5**, 727–732 (1995).
- Sirigu, A. et al. Congruent unilateral impairments for real and imagined hand movements. *Neuroreport* **6**, 997–1001 (1995).
- Sirigu, A. et al. The mental representation of hand movements after parietal cortex damage. *Science* **273**, 1564–1568 (1996).
- Papaxanthis, C., Schieppati, M., Gentili, R. & Pozzo, T. Imagined and actual arm movements have similar durations when performed under different conditions of direction and mass. *Exp. Brain Res.* **143**, 447–452 (2002).
- Buccino, G. et al. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* **13**, 400–404 (2001).
- Crammond, D. J. Motor imagery: never in your wildest dream. *Trends Neurosci.* **20**, 54–57 (1997).
- Jeannerod, M. Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage* **14**, S103–S109 (2001).
- Johnson, S. H. et al. Selective activation of a parietofrontal circuit during implicitly imagined prehension. *Neuroimage* **17**, 1693–1704 (2002).
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. Understanding motor events: a neurophysiological study. *Exp. Brain Res.* **91**, 176–180 (1992).
- Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *Int. J. Psychophysiol.* **35**, 165–177 (2000).
- Rizzolatti, G., Fogassi, L. & Gallese, V. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Rev. Neurosci.* **2**, 661–670 (2001).
- Umiltà, M. A. et al. I know what you are doing. A neurophysiological study. *Neuron* **31**, 155–165 (2001).
- Jeannerod, M., Arbib, M. A., Rizzolatti, G. & Sakata, H. Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.* **18**, 314–320 (1995).
- Ramnan, N. & Miall, R. C. A system in the human brain for predicting the actions of others. *Nature Neurosci.* **7**, 85–90 (2004).
- Flanagan, J. R. & Johansson, R. S. Action plans used in action observation. *Nature* **424**, 769–771 (2003).
- Ferrari, P. F., Gallese, V., Rizzolatti, G. & Fogassi, L. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* **17**, 1703–1714 (2003).
- Cisek, P. & Kalaska, J. F. Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex. *J. Neurophysiol.* **87**, 1149–1154 (2002).
- Watanabe, M. et al. Behavioral reactions reflecting differential reward expectations in monkeys. *Exp. Brain Res.* **140**, 511–518 (2001).
- Kohler, E. et al. Hearing sounds, understanding actions: action representation in mirror neurons. *Science* **297**, 846–848 (2002).
- Shen, L. & Alexander, G. E. Preferential representation of instructed target location versus limb trajectory in dorsal premotor area. *J. Neurophysiol.* **77**, 1195–1212 (1997).
- Ochiai, T., Mushiaki, H. & Tanji, J. Effects of image motion in the dorsal premotor cortex during planning of an arm movement. *J. Neurophysiol.* **88**, 2167–2171 (2002).
- Cisek, P., Crammond, D. J. & Kalaska, J. F. Neural activity in primary motor and dorsal premotor cortex in reaching tasks with the contralateral versus ipsilateral arm. *J. Neurophysiol.* **89**, 922–942 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Ajemian, T. Drew and S. Wise for comments and suggestions on earlier versions of this manuscript. This study was supported by operating grants to J.F.K. from the Canadian Institutes of Health Research, the Human Frontier Science Program, and the New Emerging Team Grant in Computational Neuroscience (CIHR), and by NIH and FCAR postdoctoral fellowships to P.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.F.K. (kalaskaj@physio.umontreal.ca).

Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells

Keisuke Ito^{1,2,*}, Atsushi Hirao^{1,*}, Fumio Arai¹, Sahoko Matsuoka¹, Keiyo Takubo¹, Isao Hamaguchi¹, Kana Nomiyama¹, Kentaro Hosokawa¹, Kazuhiro Sakurada³, Naomi Nakagata⁴, Yasuo Ikeda², Tak W. Mak⁵ & Toshio Suda¹

¹Department of Cell Differentiation, The Sakaguchi Laboratory of Developmental Biology, Keio University School of Medicine, 35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan

²Division of Hematology, Department of Internal Medicine, Keio University School of Medicine, 35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan

³BioFrontier Laboratories, Kyowa Hakko Kogyo Co. Ltd, Machida-shi, Tokyo 194-8533, Japan

⁴Division of Reproductive Engineering, Center for Animal Resources and Development, Kumamoto University, Honjo, Kumamoto 860-0811 Japan

⁵Advanced Medical Discovery Institute, Ontario Cancer Institute, Departments of Medical Biophysics and Immunology, University of Toronto, 620 University Avenue, Suite 706, Toronto, Ontario M5G 2C1, Canada

* These authors contributed equally to this work

The 'ataxia telangiectasia mutated' (*Atm*) gene maintains genomic stability by activating a key cell-cycle checkpoint in response to DNA damage, telomeric instability or oxidative stress^{1,2}. Mutational inactivation of the gene causes an autosomal recessive disorder, ataxia-telangiectasia, characterized by immunodeficiency, progressive cerebellar ataxia, oculocutaneous telangiectasia, defective spermatogenesis, premature ageing and a high incidence of lymphoma^{3,4}. Here we show that ATM has an essential function in the reconstitutive capacity of haematopoietic stem cells (HSCs) but is not as important for the proliferation or differentiation of progenitors, in a telomere-independent manner. *Atm*^{-/-} mice older than 24 weeks showed progressive bone marrow failure resulting from a defect in HSC function that was associated with elevated reactive oxygen species. Treatment with anti-oxidative agents restored the reconstitutive capacity of *Atm*^{-/-} HSCs, resulting in the prevention of bone marrow failure. Activation of the p16^{INK4a}-retinoblastoma (Rb) gene product pathway in response to elevated reactive oxygen species led to the failure of *Atm*^{-/-} HSCs. These results show that the self-renewal capacity of HSCs depends on ATM-mediated inhibition of oxidative stress.

It has been shown that many factors are crucial in determining the fate of stem cells, yet molecular mechanisms governing the maintenance of the self-renewal capacity have not been well studied. Wong *et al.*⁵ recently reported an interesting linkage between ATM and stem cell function. This group demonstrated that *Atm* deficiency, in combination with telomerase RNA component (*Terc*) deficiency, led to neural stem cell failure characterized by increased telomere erosion. In contrast, *Atm* deficiency alone did not affect neural stem cells. To investigate roles of ATM on self-renewal of stem cells in detail, we evaluated the effects of *Atm* deficiency on the haematopoietic system in this study.

Other than the previously reported defect in T cell development⁶⁻⁹, *Atm*^{-/-} mice showed normal numbers and subsets of haematopoietic cells in peripheral blood at 8 weeks of age. The loss of ATM did not show any defect in the differentiation or proliferation potential of progenitors at this age (Supplementary Fig. 1). The percentages of S/G2/M phases of the cell cycle in c-Kit⁺Sca-1⁺ Lineage⁻ (KSL) cells were equal in wild-type (WT; 12.15 ± 1.60%) and *Atm*^{-/-} mice (13.36 ± 2.81%). Next we investigated whether more primitive haematopoietic cells, HSCs, are affected

in *Atm*^{-/-} bone marrow (BM). The frequencies of the stem cell subsets CD34⁺lowKSL, Thy1^{low}KSL and Tie2⁺KSL (as defined by cell surface markers¹⁰⁻¹²) were significantly reduced in *Atm*^{-/-} BM (Fig. 1a). The number of colony-forming cells derived from *Atm*^{-/-} KSL cells after 6 weeks of culture on stromal cells was significantly decreased compared with that of the WT, although a short-term culture (less than 2 weeks) did not show a significant difference (Fig. 1b). To assess the repopulating ability of *Atm*^{-/-} HSCs directly *in vivo*, we performed a competitive reconstitution assay in which BM mononuclear cells (MNCs) from *Atm*^{-/-} mice competed with BM MNCs from congenic (Ly5.1/Ly5.2) WT mice to reconstitute the haematopoietic compartment of an irradiated recipient mouse (Ly5.1). Flow cytometric analysis of peripheral blood of the transplanted recipients taken at 4 weeks after transplantation revealed that *Atm*^{-/-} BM MNCs were just as able as the competitor cells to contribute to haematopoietic reconstitution (Fig. 1c, left panel). Thus, short-term repopulation was not affected. However, there were markedly fewer haematopoietic cells derived from *Atm*^{-/-} BM MNCs than from competitor MNCs at 16 weeks after transplantation (Fig. 1c, right panel), indicating that *Atm*^{-/-} HSCs lack long-term repopulating capacity. To examine this phenomenon more closely, we transplanted purified *Atm*^{-/-} KSL cells into congenic recipients. We observed a marked impairment of long-term repopulation capacity (Fig. 1d), which affected the B, T and myeloid cell lineages (data not shown). These transplantation data did not result from altered homing or apoptosis of HSCs in the absence of ATM (Supplementary Fig. 2 and 3). Thus, ATM has an essential role in the self-renewal of adult HSCs but is less important for their proliferation or differentiation into haematopoietic progenitor cells.

The capacity for self-renewal under conditions of stress such as transplantation may or may not reflect the expansion of HSCs under normal homeostatic conditions. We therefore evaluated the effects of *Atm* deficiency on haematopoiesis in older mice. Although *Atm*^{-/-} mice often die of lymphoma after 9 weeks of age⁶⁻⁹, some lymphoma-free animals do survive. When we monitored haematopoiesis in these mice at 24 weeks of age, they all exhibited a progressive anaemia not seen in the WT mice (mean haemoglobin and haematocrit values at 24 weeks were 15.3 ± 0.94 g dl⁻¹ and 50.3 ± 0.47%, respectively, in the WT (*n* = 5), in comparison with 9.93 ± 0.37 g dl⁻¹ and 28.0 ± 1.41%, respectively, in *Atm*^{-/-} mice (*n* = 5); Fig. 1e). The numbers of leukocytes and platelets in *Atm*^{-/-} mice were also significantly lower than in the WT (mean white blood cell and platelet counts were 10,800 ± 210 μl⁻¹ and 126 ± 19 × 10⁴, respectively, in the WT (*n* = 5), in comparison with 7,300 ± 940 μl⁻¹ and 76 ± 6 × 10⁴, respectively, in *Atm*^{-/-} mice (*n* = 5)). Analysis of BM from *Atm*^{-/-} mice revealed a decrease in cellularity accompanied by a relative increase in adipose tissue (Fig. 1f; the mean ratio of number of BM MNCs from *Atm*^{-/-} mice to that from WT mice was 0.37 ± 0.02). Neither cells with abnormal morphology nor lymphoma-like cells were observed in *Atm*^{-/-} BM. Flow cytometric analysis of BM revealed that the absolute cell numbers of multiple lineages, including myeloid cells (Mac-1⁺Gr-1⁺), B cells (B220⁺) and erythroid cells (Ter119⁺ cells), were decreased in *Atm*^{-/-} BM (Fig. 1g). Colony-forming assays confirmed that the numbers of myeloid and erythroid precursors among *Atm*^{-/-} BM MNCs were markedly decreased (Supplementary Fig. 4a). Co-culture of BM MNCs on stromal cells showed that *Atm*^{-/-} cells were no longer able to form any colonies after 2 weeks of culture (Supplementary Fig. 4b).

To analyse the HSC population in older *Atm*^{-/-} mice in more detail, we examined the marker expression and colony-forming ability of the KSL fraction of *Atm*^{-/-} BM cells. In contrast to the normal frequency of KSL cells observed in 8-week-old *Atm*^{-/-} mice (Supplementary Fig. 1a), the c-Kit^{high} fraction of Sca-1⁺Lineage⁻ cells had disappeared in 24-week-old *Atm*^{-/-} mice, and most c-Kit⁺Sca-1⁺Lineage⁻ cells expressed only low levels of c-Kit

(Fig. 1h). It has been shown that the c-Kit^{high} fraction in Sca-1⁺Lineage⁻ cells includes the long-term repopulating HSCs, and that c-Kit^{low} cells do not have this capacity¹³. Consistent with that report, our long-term culture of WT BM cells revealed that the c-Kit^{high} fraction, but not the c-Kit^{low} fraction, of the Sca-1⁺Lineage⁻ contained HSCs (Supplementary Fig. 4c). As expected, the c-Kit^{low}Sca-1⁺Lineage⁻ cells found in *Atm*^{-/-} mice did not form any colonies in long-term cultures. Furthermore, repopulating cells derived from *Atm*^{-/-} BM MNCs were no longer observed in recipient mice at either 4 or 16 weeks after transplantation (Supplementary Fig. 4d). Thus, our data indicate that chronic *Atm* deficiency *in vivo* results in progressive multi-lineage BM failure due to defective maintenance of the adult HSC pool.

To explain the mechanism underlying the regulation of the HSC pool by ATM, we investigated whether the defect in *Atm*^{-/-} HSCs was caused by telomere dysfunction. The telomere length of BM cells was not affected by the loss of *Atm* alone (Supplementary Fig. 5). This result is consistent with a previous report showing that telomere dysfunction (as evaluated by anaphase bridging) was not observed in *Atm*^{-/-} mice⁵. We also examined whether enhance-

ment of telomerase activity could restore the repopulating capacity of *Atm*^{-/-} HSCs. To this end, we used retroviral infection to introduce telomere reverse transcriptase (TERT) into *Atm*^{-/-} KSL cells. However, overexpression of TERT was not able to rescue the defect (Table 1). These data indicate that ATM is required for the maintenance of HSC repopulation capacity in a telomere-independent manner.

In addition to telomere maintenance, it has been suggested that ATM might be involved in oxidative defence through the induction of the major anti-oxidative systems mediated by catalase, glutathione peroxidase, superoxide dismutase and glutathione reductase¹⁴. To determine whether the altered HSC function in *Atm*^{-/-} mice was related to oxidative stress, we examined the concentration of reactive oxygen species (ROS) in HSCs. The intracellular concentration of H₂O₂, which was evaluated by 2',7'-dichlorofluorescein diacetate (DCF-DA) staining, was higher in KSL cells from 8-week-old *Atm*^{-/-} mice than from WT controls (Fig. 2a, left panel). The elevation of H₂O₂ in *Atm*^{-/-} KSL cells was even more pronounced after 2 days of culture with stem cell factor (SCF) and thrombopoietin, which support HSC cell division¹⁵

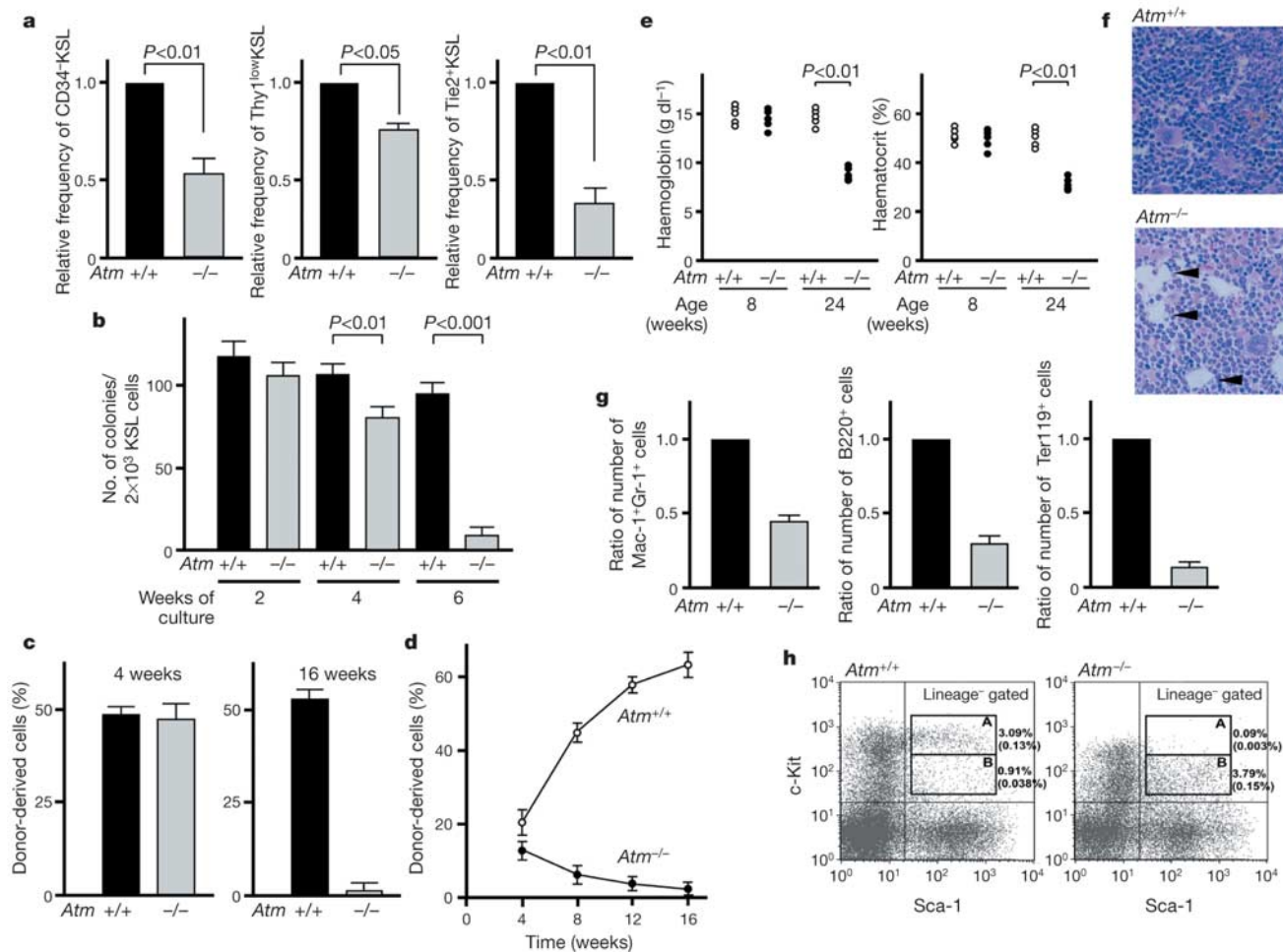


Figure 1 Defective haematopoiesis in the absence of ATM. **a**, Decreased frequencies of stem cell subsets. Results are mean relative frequencies $\pm$ s.d. for CD34⁺KSL (left), Thy1⁺KSL (middle) and Tie2⁺KSL (right) cells in *Atm*^{-/-} BM compared with *Atm*^{+/+} BM ($n = 4$). **b**, Decreased colony formation after long-term culture of *Atm*^{+/+} and *Atm*^{-/-} KSL cells (2×10^3). Results are means $\pm$ s.d. ($n = 5$). **c, d**, Defective haematopoietic reconstitution. Recipient mice were transplanted with 4×10^5 BM MNCs (**c**) or 2×10^3 KSL cells (**d**) from *Atm*^{+/+} or *Atm*^{-/-} mice in competition assays. Results are mean percentages $\pm$ s.d. of donor-derived cells ($n = 5$). **e**, Concentrations of haemoglobin (left panel) and haematocrit (right panel) in peripheral blood of individual

Atm^{+/+} and *Atm*^{-/-} mice of the indicated ages ($n = 5$). **f**, Decreased BM cellularity. Sections of BM from 24-week-old *Atm*^{+/+} and *Atm*^{-/-} mice were stained with haematoxylin and eosin. Arrows indicate adipose tissue. **g**, Decreased numbers of myeloid cells (left), B cells (middle) and erythroid cells (right) in *Atm*^{-/-} BM. Results are mean relative frequencies $\pm$ s.d. of cells in *Atm*^{-/-} BM compared with *Atm*^{+/+} BM ($n = 4$). **h**, Decreased KSL cells in 24-week-old *Atm*^{-/-} mice. For each panel, inset A shows c-Kit^{high}Sca-1⁺Lineage⁻ cells and inset B shows c-Kit^{low}Sca-1⁺Lineage⁻ cells. The frequencies of KSL cells as a percentage of total Lineage⁻ cells, and of total BM cells (in brackets) are indicated. Results are representative of four mice per genotype.

(Fig. 2a, right panel). In an effort to manipulate intracellular ROS concentrations, we treated WT and *Atm*^{-/-} KSL cells with the permeant thiol *N*-acetyl-L-cysteine (NAC), which acts as an anti-oxidative agent¹⁶. Treatment of *Atm*^{-/-} KSL cells with NAC significantly decreased the concentration of intracellular ROS (Fig. 2a, right panel).

To determine whether the defective repopulating capacity of *Atm*^{-/-} mice could be restored by decreasing their elevated intracellular ROS, we established 6-week cultures of WT and *Atm*^{-/-} KSL cells on stromal cells and treated them with one of the anti-oxidants NAC or catalase. The numbers of colonies formed from *Atm*^{-/-} HSCs were restored to near-WT levels after treatment with

either anti-oxidative agent (Fig. 2b). To replicate this phenomenon *in vivo*, we decreased the ROS in cells of 8-week-old *Atm*^{-/-} mice by the daily subcutaneous administration of NAC (100 mg kg⁻¹ d⁻¹) for 3 weeks. We found that the treatment of *Atm*^{-/-} mice with NAC before transplantation improved haematopoietic reconstitution at 16 weeks after transplantation to a degree, but that total repopulation was still significantly lower than when WT KSL were used (Fig. 2c). Neither did NAC treatment of the recipient mice after transplantation with *Atm*^{-/-} KSL cells completely rescue reconstitution. However, when NAC was administered both before and after BMT, the level of repopulation achieved was comparable to that of the WT. Treatment with NAC either before or after transplantation, or at both times, did not affect the repopulating capacity of WT HSCs in our system (Fig. 2c and data not shown). The rescued reconstitutive capacity of *Atm*^{-/-} HSCs was confirmed by the detection of donor-derived B cells, T cells and myeloid cells in recipient mice (data not shown). These results show that the elevated ROS present in *Atm*^{-/-} mice are the primary cause of the defective repopulating capacity of *Atm*^{-/-} HSCs.

We next investigated whether long-term treatment with NAC could prevent BM failure in older *Atm*^{-/-} mice. NAC was administered daily to 4-week-old WT and *Atm*^{-/-} mice until they reached the age of 24 weeks. Although H₂O₂ concentrations in *Atm*^{-/-} KSL cells were initially elevated over those in WT KSL cells, treatment with NAC decreased H₂O₂ concentrations to those in WT KSL cells

Table 1 Effects of TERT overexpression on haematopoietic reconstitution

Cell type	No. of mice showing reconstituted haematopoiesis (% of donor cells)	
	EGFP	TERT
WT	4/4 (32–46%)	4/4 (35–48%)
<i>Atm</i> ^{-/-}	0/4 (0%)	0/4 (0%)

KSL cells (5 × 10³) were incubated with SCF and thrombopoietin and infected with retrovirus carrying TERT/EGFP or EGFP. The enhanced telomerase activity with exogenous TERT was confirmed by telomeric repeat amplification protocol assay. These cells (Ly5.2) were transplanted into Ly5.1 mice along with 4 × 10⁵ Ly5.1 BM MNCs, which served as competitor cells. Haematopoietic reconstitution was confirmed by the detection of more than 1% donor-derived peripheral blood cells 16 weeks after transplantation. The infection efficiency of both retroviral vectors was 40–60% of WT and *Atm*^{-/-} KSL cells as evaluated by detection of EGFP expression.

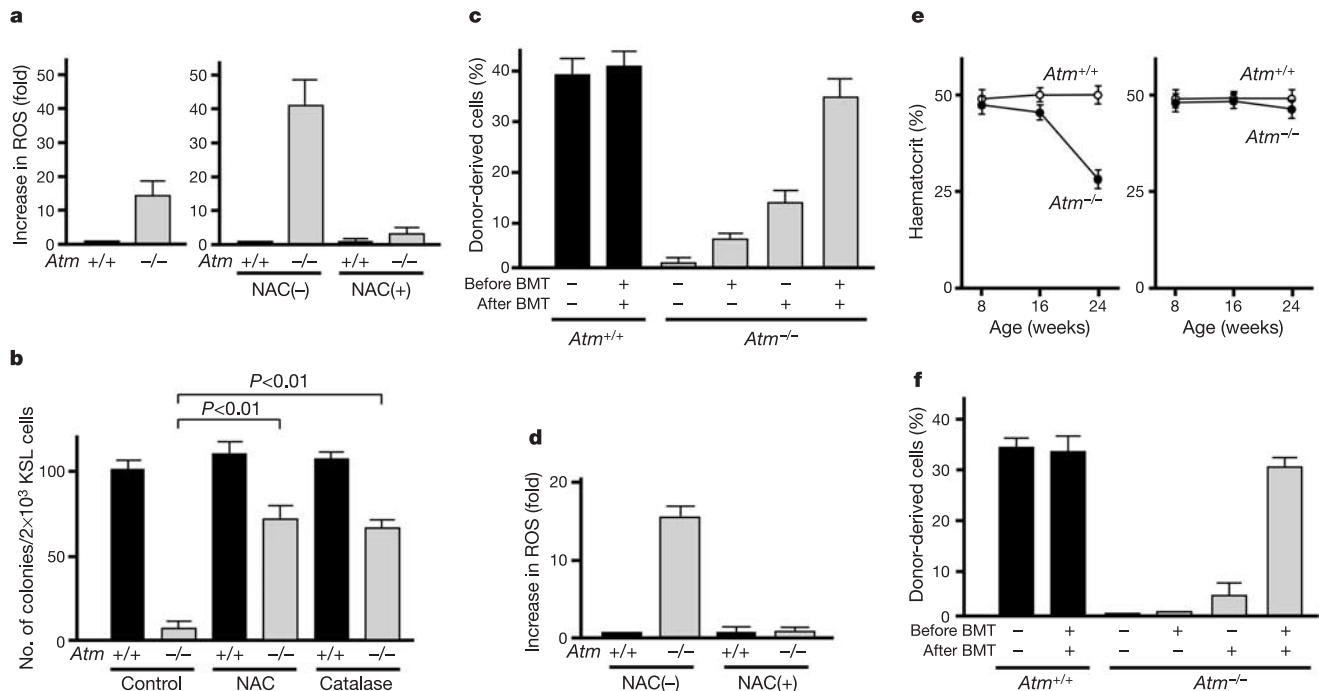


Figure 2 The defect in *Atm*^{-/-} HSC function is caused by elevated ROS. **a**, Increased intracellular ROS. Left, H₂O₂ concentrations were determined by DCF-DA staining in freshly isolated KSL cells from 8-week-old *Atm*^{+/+} and *Atm*^{-/-} mice. Right, KSL cells were also incubated with cytokines alone or with cytokines plus the anti-oxidant NAC. Results are expressed as mean fold increases ± s.d. in intracellular ROS in cells in comparison with *Atm*^{+/+} cells incubated with cytokines alone (*n* = 5). **b**, Restoration of colony formation after long-term culture with anti-oxidative agents. KSL cells were cultured for 6 weeks on OP9 stromal cells with or without 100 μM NAC or 100 U ml⁻¹ catalase. Results are means ± s.d. (*n* = 5). **c**, Restored haematopoietic reconstitution *in vivo* with NAC treatment. KSL cells (5 × 10³) were isolated from 8-week-old *Atm*^{+/+} or *Atm*^{-/-} mice that had been treated subcutaneously with NAC (100 mg kg⁻¹ d⁻¹) for 3 weeks (indicated as 'before BMT'). Some recipient mice were also treated with NAC after transplantation ('after BMT'). Results are mean percentages ± s.d. of donor-derived

cells at 16 weeks after transplantation (*n* = 3). **d**, Reduction in intracellular ROS *in vivo* with NAC treatment. H₂O₂ concentrations were determined in KSL cells from 24-week-old *Atm*^{+/+} and *Atm*^{-/-} mice that had been either left untreated or treated with NAC for 20 weeks. Results are mean fold increases ± s.d. in H₂O₂ in comparison with *Atm*^{+/+} cells without NAC (*n* = 3). **e**, Prevention of anaemia by treatment with NAC. Mean concentrations ± s.d. of haematocrit in peripheral blood of untreated (left) or NAC-treated (right) *Atm*^{+/+} and *Atm*^{-/-} mice of the indicated ages are shown (*n* = 4). **f**, Restored haematopoietic reconstitution *in vivo* with long-term treatment with NAC. KSL cells (5 × 10³) were isolated from 24-week-old *Atm*^{+/+} or *Atm*^{-/-} mice that had been treated with NAC for 20 weeks (indicated as 'before BMT'). Some recipient mice were also treated with NAC after transplantation ('after BMT'). Results are mean percentages ± s.d. of donor-derived cells at 16 weeks after transplantation (*n* = 3).

from 24-week-old $Atm^{-/-}$ mice, as expected (Fig. 2d). $Atm^{-/-}$ mice that received long-term treatment with NAC did not show anaemia (Fig. 2e), a decrease in progenitor colony-forming capacity (Supplementary Fig. 6a), or reduced frequencies of stem cell subsets (Supplementary Fig. 6b, c). We therefore concluded that the observed decrease in ROS prevented BM failure in older $Atm^{-/-}$ mice. Although the HSCs of older $Atm^{-/-}$ mice lose their repopulating capacity even when NAC is given to recipient mice after transplantation, a combination of long-term treatment with NAC both before and after transplantation completely rescued stem cell function (Fig. 2f). These results show that it is the elevation in ROS, and not some other factor, that causes the defect in $Atm^{-/-}$ stem cell function *in vivo*.

To understand better how the elevation in ROS affects $Atm^{-/-}$ HSCs, we used polymerase chain reaction with reverse transcription (RT-PCR) to compare the gene expression profiles of KSL cells freshly isolated from the BM of 8-week-old WT and $Atm^{-/-}$ mice. Several genes involved in the maintenance of the HSC compartment were analysed, including *HOXB4* (ref. 17), *Bmi-1* (refs 18, 19) and those encoding CDK inhibitors^{20,21}. However, no obvious differences in gene expression were observed between WT and $Atm^{-/-}$ mice (Fig. 3a). We next attempted to mimic transplantation conditions by incubating KSL cells with cytokines *in vitro* for 2 days, followed by gene expression profiling. The tumour suppressors p16^{INK4a} and p19^{ARF}, which are generated by alternative

splicing from the *INK4a* locus²², were highly elevated in $Atm^{-/-}$ KSL cells under these conditions (Fig. 3b). In contrast, p21^{WAF1/CIP1} and p27^{KIP1}, which are members of a different CDK inhibitor family, were not affected (Fig. 3b and data not shown). Treatment with NAC abrogated the upregulation of p16^{INK4a} and p19^{ARF} expression in $Atm^{-/-}$ KSL cells *in vitro* (Fig. 3b), indicating that the elevation of ROS that occurs in the absence of *Atm* controls p16^{INK4a} and p19^{ARF} expression as well as repopulating capacity. To investigate whether p16^{INK4a} and p19^{ARF} were similarly affected *in vivo*, we analysed gene expression profiles of immature haematopoietic cell populations (the KSL and Lineage⁻ fractions) in 24-week-old $Atm^{-/-}$ mice with BM failure. The expression levels of both p16^{INK4a} and p19^{ARF} in $Atm^{-/-}$ cells were significantly higher than those in the WT (Fig. 3c and Supplementary Fig. 7). Furthermore, long-term treatment with NAC decreased the expression of p16^{INK4a} and p19^{ARF} in 24-week-old $Atm^{-/-}$ KSL cells *in vivo* (Fig. 3d and Supplementary Fig. 7). Upregulation of p16^{INK4a} and p19^{ARF} by the loss of Bmi-1, a transcriptional repressor of the p16^{INK4a} and p19^{ARF} genes²³, led to defective self-renewal of adult HSCs and neural stem cells, but was less critical for the generation of the differentiated progeny of these cell types^{18,19,24}. We therefore speculated that the increased ROS present in $Atm^{-/-}$ mice might cause the loss of the HSC pool through the upregulation of p16^{INK4a} and p19^{ARF}.

To test this hypothesis, we reduced p16^{INK4a} and p19^{ARF}

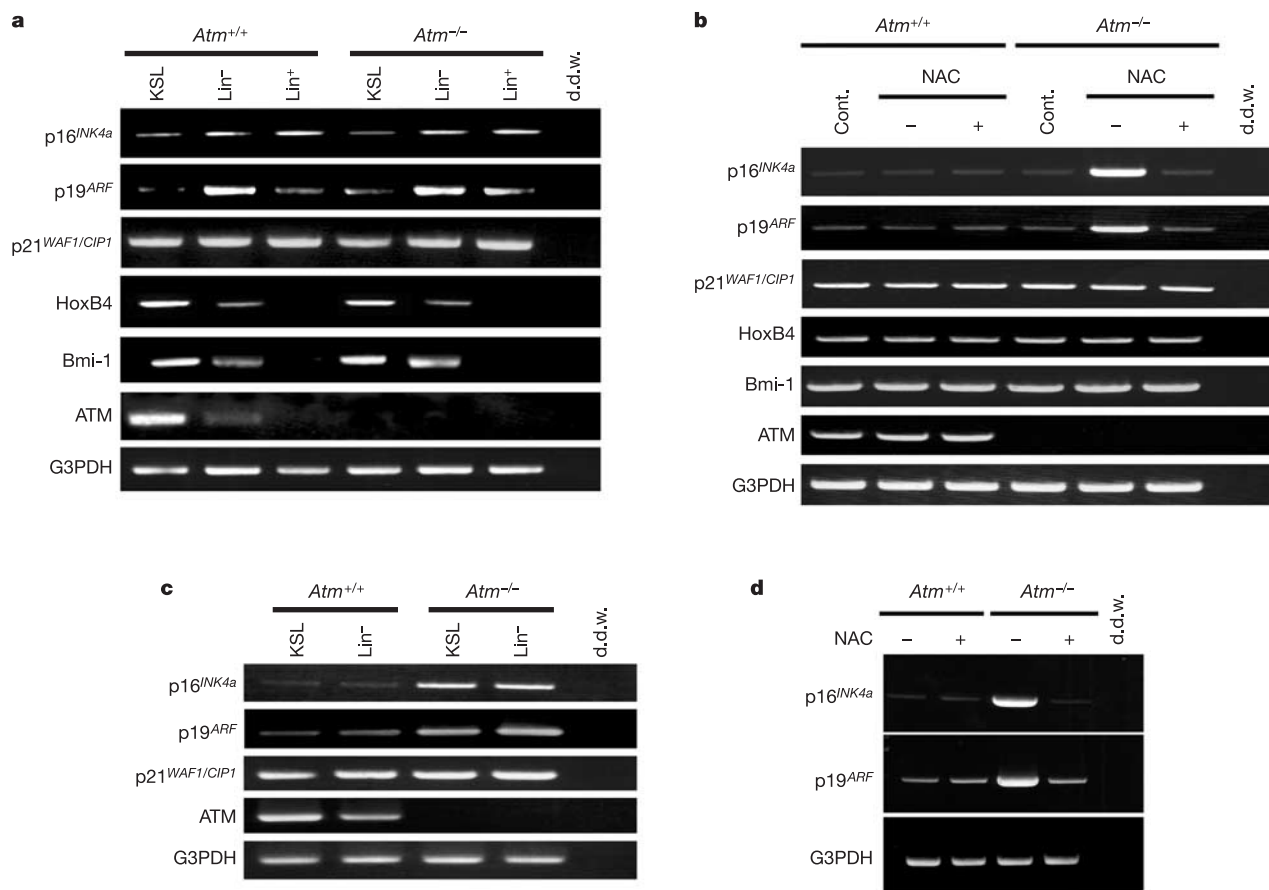


Figure 3 Elevated ROS induces upregulation of p16^{INK4a} and p19^{ARF} in $Atm^{-/-}$ KSL cells. **a**, Normal gene expression profile in haematopoietic fractions of 8-week-old $Atm^{-/-}$ mice. One result representative of at least three trials is shown. Lin, Lineage; d.d.w., double-distilled water. **b**, Upregulation of p16^{INK4a} and p19^{ARF} expression in cultured $Atm^{-/-}$ KSL cells *in vitro*. KSL cells were obtained from 8-week-old $Atm^{+/+}$ and $Atm^{-/-}$ mice, and gene expression profiles were determined in KSL cells that had been either freshly isolated (Cont.), incubated with cytokines alone (NAC⁻) or incubated

with cytokines plus NAC (NAC⁺). **c**, Upregulation of p16^{INK4a} and p19^{ARF} expression in KSL cells from aged $Atm^{-/-}$ mice. Gene expression profiles of KSL cells from 24-week-old $Atm^{+/+}$ and $Atm^{-/-}$ mice are shown. **d**, Restoration of normal p16^{INK4a} and p19^{ARF} expression *in vivo* by NAC treatment. Gene expression profiles were determined in KSL cells from 24-week-old $Atm^{-/-}$ mice that had been either left untreated or treated with NAC for 20 weeks. G3PDH, glyceraldehyde-3-phosphate dehydrogenase.

expression in KSL cells by the retroviral introduction of *Bmi-1*. Normalization of $p16^{INK4a}$ and $p19^{ARF}$ expression by the overexpression of *Bmi-1* was confirmed by RT-PCR (Fig. 4a). Transplantation of WT KSL cells expressing either *Bmi-1* or the control vector enhanced green fluorescent protein (EGFP) reconstituted donor-derived haematopoiesis in recipient mice. However, transplanted *Atm*^{-/-} KSL cells could restore HSC repopulating capacity in the recipient only when the cells were infected with the *Bmi-1*-expressing retrovirus and not with the control EGFP-expressing retrovirus (Fig. 4b). These results indicated that upregulation of either $p16^{INK4a}$ or $p19^{ARF}$ causes the defective stem cell function observed in *Atm*^{-/-} mice. It has been shown that upregulation of $p16^{INK4a}$ inhibits the phosphorylation of Rb family members (that is, it inhibits the inactivation of Rb function), whereas upregulation of $p19^{ARF}$ inhibits the degradation of p53 protein²². To identify which pathway, $p16^{INK4a}$ -Rb or $p19^{ARF}$ -p53, governs stem cell function, we inhibited the function of Rb family members or p53 by infecting *Atm*^{-/-} KSL cells with a retroviral vector expressing human papilloma virus type 16 E7 and E6, respectively²⁵. The overexpression of E7, but not E6, restored the repopulating capacity of *Atm*^{-/-} HSCs, indicating that it is the $p16^{INK4a}$ -Rb pathway that contributes to defective stem cell function in response to the elevated ROS in *Atm*^{-/-} HSCs (Fig. 4c).

Haematopoietic progenitors have a high replicative potential and express telomerase to protect the ends of their chromosomes. As an animal ages, the telomeres undergo erosion. HSCs can also undergo telomere shortening, as reported in an examination of the serial transplantation of HSCs²⁶. Accelerated telomere erosion due to the loss of *Terc* resulted in a decrease in the long-term repopulating capacity of these HSCs^{27,28}. However, overexpression of *TERT* in

mouse HSCs to prevent telomere erosion has no effect on the longevity of these cells²⁹, indicating that telomere-independent mechanisms might regulate the self-renewal of HSCs. Indeed, our study supports a model in which HSC self-renewal is regulated by ROS, which are in turn controlled by ATM (Fig. 4d). Although there are discrepancies between the phenotypes of patients with ataxia-telangiectasia and *Atm*^{-/-} mice, we believe that our results shed light on a possible ATM function on the self-renewal of human stem cells and that the defect in HSC capacity might link to ataxia-telangiectasia pathophysiology. The recovery of HSC function by treatment with NAC indicates that patients with ataxia-telangiectasia might benefit from the administration of anti-oxidative agents as a pharmaceutical treatment. Our work could have repercussions outside the field of haematology, because the signalling pathways regulating stem cell self-renewal might also be used by cancer cells to drive malignant proliferation³⁰. The idea leads to the intriguing notion that tumour suppressors might participate in the regulation of stem cell function. It supports a model in which master regulator molecules govern the disparate processes of stem cell self-renewal, normal ageing and tumour development. □

Methods

Mice

Atm^{+/-} mice were a gift from P. J. McKinnon (St Jude Children's Research Hospital, Memphis, Tennessee). Offspring of these mice were genotyped by PCR-based assays with mouse tail DNA. Littermates were used as controls in all experiments. C57BL/6 mice congenic for the *Ly5* locus (B6-Ly5.1) were purchased from Sankyo-Lab Service.

Long-term cultures and colony-forming assays

For long-term cultures, BM MNCs or KSL cells were co-cultured with OP9 cells in MEM (Sigma) containing 12.5% fetal calf serum (JRH Bioscience), 12.5% horse serum (Gibco BRL) and 1 nM dexamethasone. After 2, 4 or 6 weeks of culture, cells were harvested and used for haematopoietic colony-forming assays. Methylcellulose colony-forming assays were performed with 20 ng ml⁻¹ SCF (PeproTech EC Ltd), 20 ng ml⁻¹ interleukin-3 (PeproTech EC Ltd), 2 U ml⁻¹ erythropoietin (Chugai Pharmaceutical Co., Ltd) as described previously¹². For some experiments, 100 μM NAC (Sigma) or 100 U ml⁻¹ catalase (Sigma) was added to long-term cultures.

Flow cytometry

The following monoclonal antibodies (mAbs) were used for flow cytometric analyses: mAbs against c-Kit (2B8), Sca-1 (E13-161.7), CD4 (L3T4), CD8 (53-6.72), B220 (RA3-6B2), TER-119, Gr-1 (RB6-8C5), CD34 (RAM34), Thy1 (53-2.1), anti-Mac-1 (M1/70) and Tie2 (TEK4). All mAbs were purchased from BD. A mixture of mAbs against CD4, CD8, B220, TER-119, Mac-1 and Gr-1 was used as a lineage marker (Lineage).

Competitive reconstitution

Lethally irradiated C57BL/6-Ly5.1 congenic mice were competitively reconstituted with 4×10^5 BM MNCs or KSL cells from *Atm*^{+/-} or *Atm*^{-/-} mice (Ly5.2), in competition with 4×10^5 BM MNCs from C57BL/6-Ly5.1/5.2 mice. Reconstitution of donor (Ly5.2) myeloid and lymphoid cells was monitored by staining blood cells with antibodies against Ly5.2, Ly5.1, CD3, B220, Mac-1 and Gr-1.

Retroviral transduction

Retroviral gene transductions of cDNAs encoding mouse *TERT*, *Bmi-1*, *E6* or *E7* were performed with the retroviral vector pMY/IRES/EGFP. KSL cells (5×10^3) from *Atm*^{+/-} or *Atm*^{-/-} mice were incubated for 24 h with 100 ng ml⁻¹ SCF and 100 ng ml⁻¹ thrombopoietin followed by incubation for 48 h with retrovirus. The infected KSL cells were transplanted into lethally irradiated Ly5.1 mice along with 4×10^5 competitor BM MNCs. The efficiency of gene transduction was evaluated by expression of EGFP.

Analysis of intracellular ROS

NAC was dissolved in saline solution at pH 7.0. For experiments *in vivo*, mice were injected subcutaneously once daily with either NAC (100 mg kg⁻¹) or saline solution. For experiments *in vitro*, cell cultures were treated with 100 μM NAC. To assess the generation of ROS in KSL cells, freshly isolated KSL cells (5×10^3) were incubated for 2 days either with cytokines (100 ng ml⁻¹ SCF and 100 ng ml⁻¹ thrombopoietin) alone, or with cytokines plus 100 μM NAC. Samples of the cultures were loaded with 5 μM DCF-DA (Sigma) and incubated on a shaker at 37 °C for 30 min. The peak excitation wavelength for oxidized DCF was 488 nm and that for emission was 525 nm.

RT-PCR

cDNAs were reverse-transcribed from total RNA prepared from 5×10^3 KSL, Lineage⁻ or Lineage⁺ cells. For some experiments, cDNA was prepared from KSL cells that had been freshly isolated and incubated for 2 days with cytokines (100 ng ml⁻¹ SCF and 100 ng ml⁻¹ thrombopoietin) alone, or with cytokines plus 100 μM NAC. Primer sequences are shown in Supplementary Table 1.

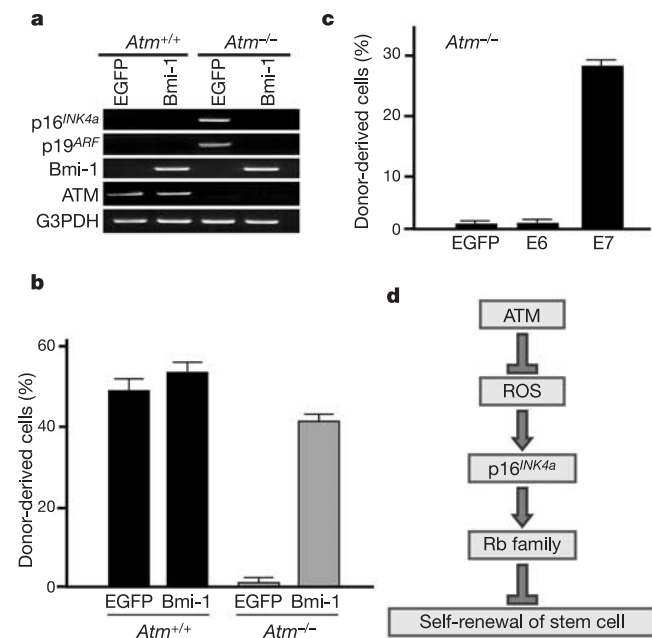


Figure 4 Activation of the $p16^{INK4a}$ -Rb pathway causes the defect in stem cell function in *Atm*^{-/-} mice. **a**, **b**, Restored haematopoietic reconstitution *in vivo* by overexpression of *Bmi-1*. KSL cells (5×10^3) isolated from 8-week-old *Atm*^{+/+} or *Atm*^{-/-} mice were infected with retrovirus carrying *Bmi-1*/EGFP or EGFP alone. The expression of $p16^{INK4a}$ and $p19^{ARF}$ in infected EGFP⁺ KSL cells collected by flow cytometry were evaluated by RT-PCR (**a**). **b**, Recipient mice were transplanted with the *Atm*^{+/+} or *Atm*^{-/-} KSL cells from **a** in competition assays. Results are mean percentages $\pm$ s.d. of donor-derived cells at 16 weeks after transplantation ($n = 3$). **c**, Restored haematopoietic reconstitution *in vivo* by overexpression of E7. Recipient mice were transplanted in competition assays with *Atm*^{-/-} KSL cells infected with retrovirus carrying either E6, E7 or EGFP as shown in **b** ($n = 3$). **d**, Proposed model for a role of ATM in the maintenance of HSC self-renewal.

Statistical analysis

P values were calculated by unpaired Student's *t*-test.

Received 13 July; accepted 30 August 2004; doi:10.1038/nature02989.

1. Meyn, M. S. Ataxia-telangiectasia and cellular responses to DNA damage. *Cancer Res.* **55**, 5991–6001 (1995).
2. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3**, 155–168 (2003).
3. Shiloh, Y. & Rotman, G. Ataxia-telangiectasia and the ATM gene: linking neurodegeneration, immunodeficiency, and cancer to cell cycle checkpoints. *J. Clin. Immunol.* **16**, 254–260 (1996).
4. Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
5. Wong, K. K. *et al.* Telomere dysfunction and Atm deficiency compromises organ homeostasis and accelerates ageing. *Nature* **421**, 643–648 (2003).
6. Barlow, C. *et al.* Atm-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159–171 (1996).
7. Xu, Y. *et al.* Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. *Genes Dev.* **10**, 2411–2422 (1996).
8. Elson, A. *et al.* Pleiotropic defects in ataxia-telangiectasia protein-deficient mice. *Proc. Natl Acad. Sci. USA* **93**, 13084–13089 (1996).
9. Herzog, K. H., Chong, M. J., Kapsetaki, M., Morgan, J. I. & McKinnon, P. J. Requirement for Atm in ionizing radiation-induced cell death in the developing central nervous system. *Science* **280**, 1089–1091 (1998).
10. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
11. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
12. Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161 (2004).
13. Randall, T. D. & Weissman, I. L. Phenotypic and functional changes induced at the clonal level in hematopoietic stem cells after 5-fluorouracil treatment. *Blood* **89**, 3596–3606 (1997).
14. Barzilai, A., Rotman, G. & Shiloh, Y. ATM deficiency and oxidative stress: a new dimension of defective response to DNA damage. *DNA Repair (Amst.)* **1**, 3–25 (2002).
15. Ema, H., Takano, H., Sudo, K. & Nakauchi, H. *In vitro* self-renewal division of hematopoietic stem cells. *J. Exp. Med.* **192**, 1281–1288 (2000).
16. Yan, M. *et al.* The ataxia-telangiectasia gene product may modulate DNA turnover and control cell fate by regulating cellular redox in lymphocytes. *FASEB J.* **15**, 1132–1138 (2001).
17. Antonchuk, J., Sauvageau, G. & Humphries, R. K. HOXB4-induced expansion of adult hematopoietic stem cells *ex vivo*. *Cell* **109**, 39–45 (2002).
18. Park, I. K. *et al.* Bmi-1 is required for maintenance of adult self-renewing hematopoietic stem cells. *Nature* **423**, 302–305 (2003).
19. Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255–260 (2003).
20. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science* **287**, 1804–1808 (2000).
21. Yuan, Y., Shen, H., Franklin, D. S., Scadden, D. T. & Cheng, T. *In vivo* self-renewing divisions of hematopoietic stem cells are increased in the absence of the early G1-phase inhibitor, p18INK4C. *Nature Cell Biol.* **6**, 436–442 (2004).
22. Lowe, S. W. & Sherr, C. J. Tumor suppression by Ink4a-Arf: progress and puzzles. *Curr. Opin. Genet. Dev.* **13**, 77–83 (2003).
23. Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* **397**, 164–168 (1999).
24. Molofsky, A. V. *et al.* Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
25. Munger, K. & Howley, P. M. Human papillomavirus immortalization and transformation functions. *Virus Res.* **89**, 213–228 (2002).
26. Allsopp, R. C., Cheshier, S. & Weissman, I. L. Telomere shortening accompanies increased cell cycle activity during serial transplantation of hematopoietic stem cells. *J. Exp. Med.* **193**, 917–924 (2001).
27. Allsopp, R. C., Morin, G. B., DePinho, R., Harley, C. B. & Weissman, I. L. Telomerase is required to slow telomere shortening and extend replicative lifespan of HSCs during serial transplantation. *Blood* **102**, 517–520 (2003).
28. Samper, E. *et al.* Long-term repopulating ability of telomerase-deficient murine hematopoietic stem cells. *Blood* **99**, 2767–2775 (2002).
29. Allsopp, R. C. *et al.* Effect of TERT over-expression on the long-term transplantation capacity of hematopoietic stem cells. *Nature Med.* **9**, 369–371 (2003).
30. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Saya for helpful discussions, T. Kiyono for providing E6 and E7 cDNA, A. Iwama for providing mouse *Bmi-1* cDNA and virus, F. Ishikawa for providing mouse TERT cDNA, T. Kitamura for providing retroviral vector pMY, A. Ono for technical support, and M. Saunders for scientific editing. A.H. was supported by grants-in-aid from the Cancer Research and from the Stem Cell Research of the Ministry of Education, Science, Sports, and Culture, Japan. T.S. was supported by a grant-in-aid from the Research for the Future Program and the Specially Promoted Research of the Ministry of Education, Science, Sports, and Culture, Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.H. (ahirao@sc.itc.keio.ac.jp) or T.S. (sudato@sc.itc.keio.ac.jp).

Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells

Hanno Hock^{1,2,3}, Melanie J. Hamblen^{1,4}, Heather M. Rooke^{1,4}, Jeffrey W. Schindler¹, Shireen Saleque¹, Yuko Fujiwara^{1,4} & Stuart H. Orkin^{1,3,4}

¹Division of Hematology/Oncology, Children's Hospital, ²Department of Medical Oncology and ³Department of Pediatric Oncology, Dana Farber Cancer Institute, Harvard Medical School and the ⁴Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

Hematopoietic stem cells (HSCs) sustain blood production throughout life. HSCs are capable of extensive proliferative expansion, as a single HSC may reconstitute lethally irradiated hosts¹. In steady-state, HSCs remain largely quiescent and self-renew at a constant low rate, forestalling their exhaustion during adult life^{2,3}. Whereas nuclear regulatory factors promoting proliferative programmes of HSCs *in vivo* and *ex vivo* have been identified^{4–6}, transcription factors restricting their cycling have

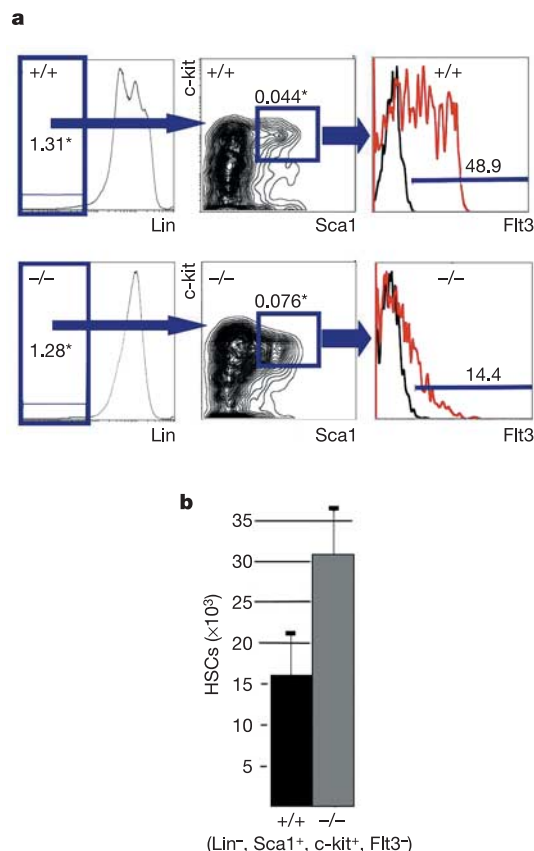
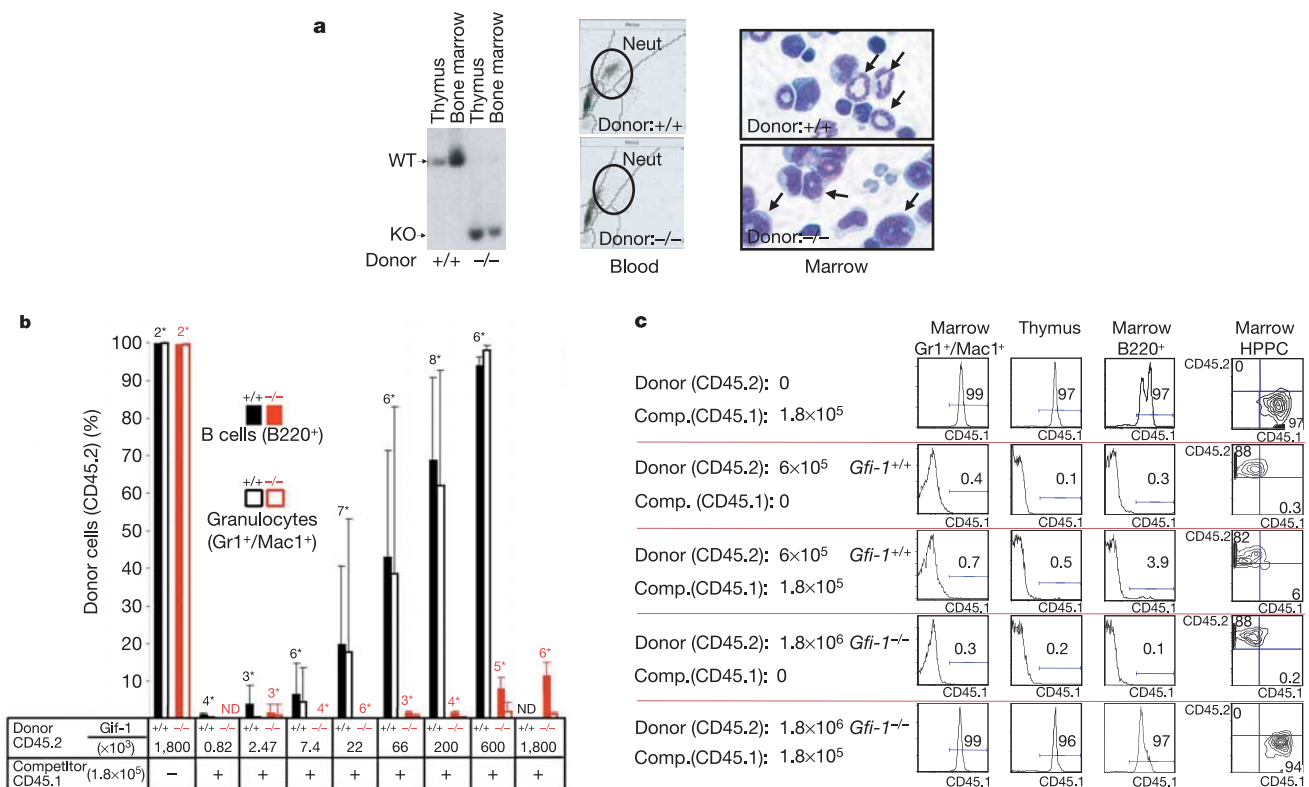


Figure 1 Immunophenotypic analysis of HSCs in *Gfi-1*^{-/-} mice. **a**, Frequency of Lin⁺, Sca1⁺, c-kit⁺ cells is elevated in *Gfi-1*^{-/-} (-/-) mice compared with wild-type littermates (+/+). Left plots show lineage gates; standard deviation (s.d.): +/+ = 0.185, -/- = 0.16. Asterisks indicate percentage of bone marrow. Middle plots identify Lin⁺, Sca1⁺, c-kit⁺ cells; s.d.: +/+ = 0.0045, -/- = 0.0016. Right plots demonstrate that Flt3 expression (red; isotype-matched control is black) within Lin⁺, Sca1⁺, c-kit⁺ cells is reduced in *Gfi-1*^{-/-} mice; s.d.: +/+ = 3.8, -/- = 2.9. Plots from one representative animal of four are shown; numbers are averages (n = 4). **b**, Total number of long-term repopulating HSCs, as average per animal (two tibia and fibula) and s.d. (n = 4, age ~4 weeks).

To address the role of Gfi-1 in HSCs we first determined the immunophenotype and HSC number in *Gfi-1*^{-/-} mice¹³ compared with wild-type littermates. The marker profile of *Gfi-1*^{-/-} HSCs was preserved, albeit with a slight but consistent decrease in the intensity of c-kit expression (Fig. 1a). We confirmed that the Lin⁻, Sca1⁺, c-kit⁺ population of both *Gfi-1*^{-/-} and control mice was similarly enriched (>250-fold) in high proliferative potential colony progenitors (HPPCs), which correlate with stem cell activity¹⁴, and that *Gfi-1*^{-/-} Lin⁻, Sca1⁺, c-kit⁺, Flt3⁻ bone marrow cells are capable of lymphomyeloid reconstitution after injection into lethally irradiated recipients (Supplementary Fig. 1). Notably, the proportion of Flt3⁺ cells within HSCs (Lin⁻, Sca1⁺, c-kit⁺) was markedly reduced (Fig. 1a). In this population Flt3 expression is associated with increased lymphoid repopulating potential but loss of long-term repopulating potential^{15,16}. Reduced Flt3 expression in this compartment may signify an early lesion in lymphoid development as *Gfi-1*^{-/-} mice have reduced lymphopoiesis^{13,17}. Surprisingly, however, the total number of phenotypic, long-term repopulating HSCs (Lin⁻, Sca1⁺, c-kit⁺, Flt3⁻) was increased in *Gfi-1*^{-/-} bone marrow (Fig. 1b). This appeared to be in contrast to findings in *Bmi-1*^{-/-} mice, where phenotypic HSCs



analysis of CD45-isotypes on blood B cells and Gr1/Mac1⁺ myeloid cells 3 months after transplantation. Bone marrow cells were given in serial dilutions at indicated numbers per recipient after lethal irradiation. Bars represent averages on groups of mice; asterisks indicate numbers of mice per group; error bars are s.d.; ND, not done. **c.** *Gfi-1*^{-/-} donor cells fail to populate haematopoietic tissues in the presence of a low dose of competitor cells. FACS analysis of CD45-isotype variants on bone marrow myeloid cells, thymocytes, bone marrow B cells or cells derived from HPPCs from representative mice of the experiment shown in **b**. Note that the profiles from the bottom row (where mice had received *Gfi-1*^{-/-} marrow (CD45.2) and competitor cells (CD45.1)) resemble the profiles from the first row where mice had received competitor only. Analysis at 4 and 7 weeks after transplantation yielded indistinguishable results (data not shown).

are markedly reduced in frequency (10–45-fold)⁴, and suggested that despite assignment of *Gfi-1* to the same complementation group as *Bmi-1* based on cooperation in lymphomagenesis, its requirement in HSCs might be distinct.

To assess the functional properties of *Gfi-1*^{-/-} HSCs we transplanted bone marrow from *Gfi-1*^{-/-} mice or littermate controls into irradiated host animals. In the absence of competitor marrow, a large dose (1.8×10^6) of *Gfi-1*^{-/-} bone marrow cells rescued recipients from lethal irradiation and sustained haematopoiesis for >3 months. The reported haematological abnormalities of *Gfi-1*^{-/-} mice—namely absence of mature granulocytes and the presence of immature myeloid cells—were evident in these reconstituted mice (Fig. 2a). Eventually, such rescued recipients became pancytopenic, succumbed to infections, or residual host haematopoiesis entirely out-competed *Gfi-1*^{-/-} cells (data not shown). Notably, *Gfi-1*^{-/-} bone marrow cells displayed a marked lack of activity in serial transplantation: transfer of 5×10^6 bone marrow cells collected from transplant recipients 6 weeks after primary transplantation of 5×10^6 bone marrow cells resulted in complete rescue and establishment of long-term haematopoiesis in lethally irradiated secondary recipients when wild-type donor cells were used. However, secondary recipients of the same dose of *Gfi-1*^{-/-} bone marrow cells succumbed to severe anaemia after 6–8 weeks and failed to display evidence of lymphomyeloid engraftment (data not shown).

To assess HSC functions in a quantitative setting, competitive repopulation experiments were performed with limiting dilutions of bone marrow from *Gfi-1*^{-/-} mice or wild-type littermates (both CD45.2) transplanted into lethally irradiated recipients along with a constant, small dose (1.8×10^5) of competitor marrow (CD45.1). Three months after transplantation lymphomyeloid reconstitution was assessed in peripheral blood B cells and granulocytes (Fig. 2b). In groups of mice that had received $\geq 7,400$ wild-type cells, CD45.2⁺ (donor) B cells and granulocytes were readily detected.

When 6×10^5 wild-type donor cells were administered, engraftment of B cells and granulocytes from competitor marrow was undetectable. In marked contrast, even 1.8×10^6 *Gfi-1*^{-/-} donor cells failed to contribute to myelopoiesis and provided only minimal contribution to B cells. These data indicate that the reconstitution of lymphomyeloid haematopoiesis by *Gfi-1*^{-/-} HSCs is reduced at least 200-fold compared with controls. Analysis of representative mice confirmed these findings in bone marrow granulocytes, thymic T cells and bone marrow B cells (Fig. 2c). Similarly, haemoglobin isoform analysis confirmed that *Gfi-1*^{-/-} donor cells provided donor-derived haemoglobin in the absence, but not the presence, of competitor marrow (data not shown). As erythroid cells do not express *Gfi-1* (ref. 13), these findings further support the presumption that competition occurred at an early level of haematopoietic development. To corroborate that competition occurred at the stem cell level and did not reflect combined effects in the differentiated lineages, we performed CD45-isotype analysis on cells recovered from HPPCs, which are believed to arise from HSCs or their early progeny¹⁴ (Fig. 2c). Again, even in mice that had received 1.8×10^6 *Gfi-1*^{-/-} cells, HPPCs were exclusively of competitor origin (Fig. 2c). Thus, competitive repopulation bone marrow transplant assays demonstrate that normal HSC activity is critically dependent on *Gfi-1*.

As young *Gfi-1*^{-/-} mice are not anaemic and have sustained myelopoiesis, we infer that *Gfi-1* is dispensable for the emergence and embryonic expansion of HSCs¹³. This conclusion is consistent with phenotypic enumeration of *Gfi-1*^{-/-} HSCs. Later in life, marrow function deteriorates in *Gfi-1*^{-/-} animals, at which time HSC function is difficult to assess because of concomitant myeloid abnormalities (ref. 13 and data not shown). To assess HSCs without prior transplantation and in the context of otherwise normal haematopoiesis, we generated chimaeric mice by injection of *Gfi-1*^{-/-} or *Gfi-1*^{+/-} (control) embryonic stem (ES) cells into wild-type blastocysts (Fig. 3a, b). In young chimaeras (<3 weeks of

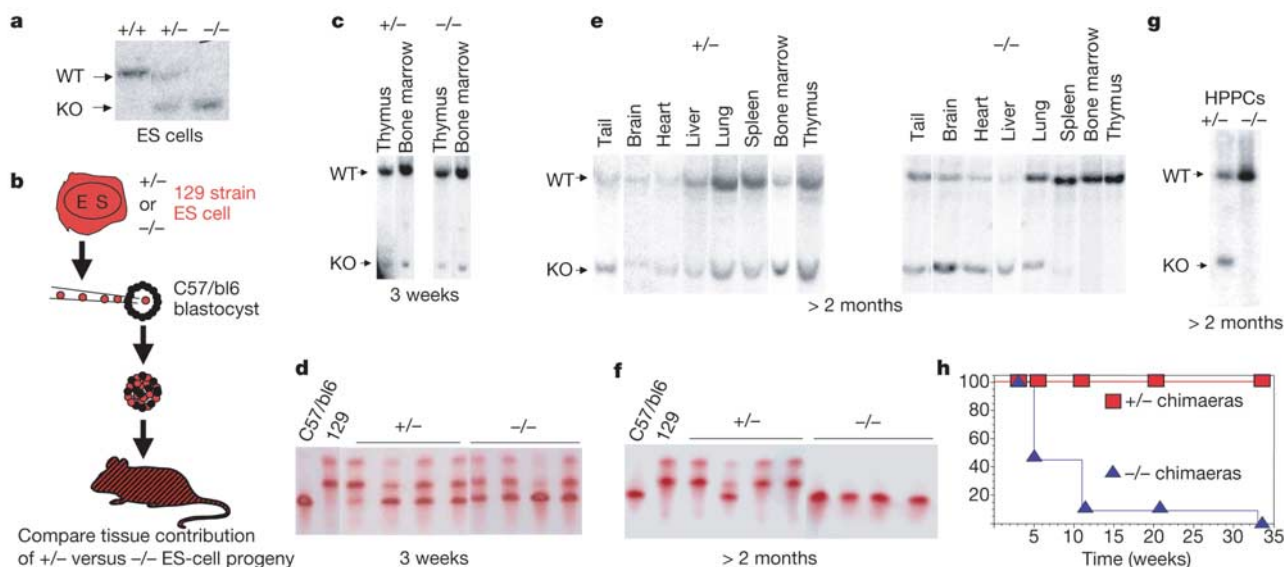


Figure 3 *Gfi-1*^{-/-} HSCs initiate, but do not sustain, bone marrow haematopoiesis in chimaeric mice. **a**, Southern blot analysis demonstrating wild-type *Gfi-1* (+/+), heterozygous *Gfi-1* disruption (+/-), and bi-allelic disruption of *Gfi-1* (-/-) in ES cell clones. **b**, Scheme of experimental design. **c**, DNA from progeny of *Gfi-1*^{+/-} or *Gfi-1*^{-/-} ES cells was detected by Southern blot analysis of haematopoietic tissues from chimaeric mice at 3 weeks of age. **d**, Haemoglobin isoform analysis demonstrating ES-cell-derived haemoglobin in the peripheral blood of chimaeras from *Gfi-1*^{-/-} and *Gfi-1*^{+/-} ES cells at 3 weeks of age. **e**, Southern blot analysis of *Gfi-1*^{+/-} and *Gfi-1*^{-/-} ES cell contribution to tissues of older chimaeras. One representative of four pairs of chimaeras

~2–3 months of age is shown. Note the specific absence of contribution to bone marrow and thymus in *Gfi-1*^{-/-} ES cell chimaeras. **f**, Haemoglobin isoform analysis demonstrating loss of contribution of *Gfi-1*^{-/-} ($n = 4$) but not *Gfi-1*^{+/-} ($n = 4$) ES cells to haemoglobin production at 2 months of age. **g**, Contribution of *Gfi-1*^{+/-} but not *Gfi-1*^{-/-} to DNA isolated from HPPCs (Southern blot). **h**, ES-cell-derived haemoglobin isoforms in peripheral blood over time. Contribution is stable in chimaeras from *Gfi-1*^{+/-} ES cells ($n = 8$ –13) but uniformly fades in chimaeras from *Gfi-1*^{-/-} ES cells ($n = 12$). The y axis shows the percentage of chimaeras with detectable ES-cell-derived haemoglobin isoforms.

age) progeny of *Gfi-1*^{+/-} and *Gfi-1*^{-/-} ES cells were detected in the bone marrow and thymus by Southern blot analysis (Fig. 3c), and ES-cell-derived haemoglobin was observed in peripheral blood (Fig. 3d). In contrast, in chimaeras >2 months, DNA from *Gfi-1*^{-/-} cells was no longer detected by Southern blot analysis in bone marrow or thymus, and was sharply reduced in spleen,

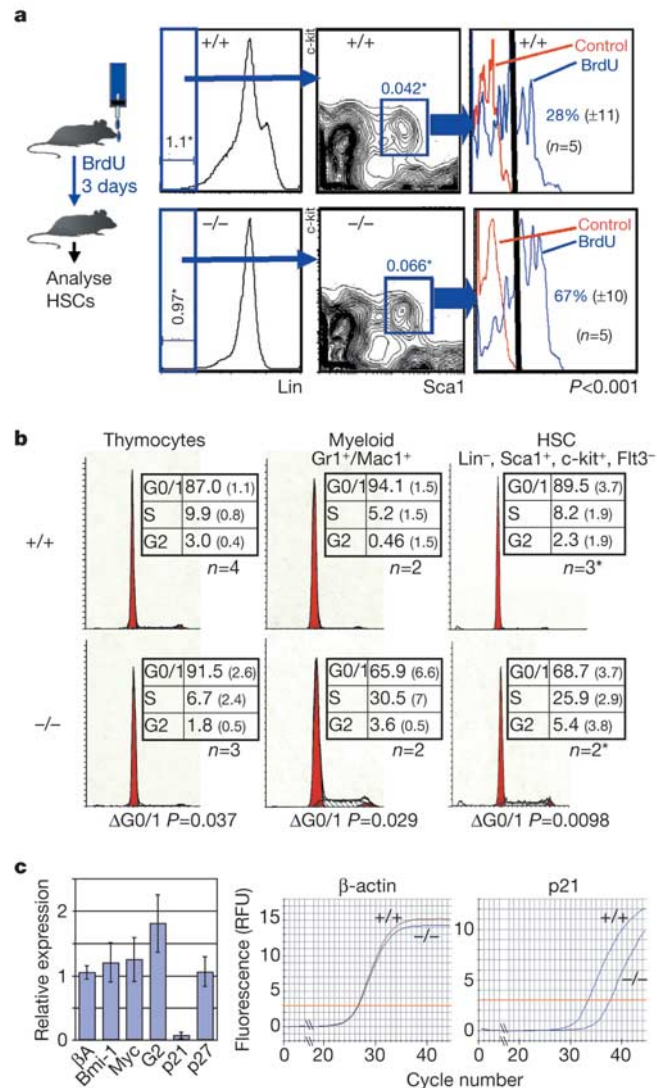


Figure 4 *Gfi-1* restricts proliferation of HSCs in a cell-context-specific manner. **a**, Analysis of bone marrow cells from animals exposed to BrdU for 3 days reveals a markedly increased fraction of HSCs that have incorporated BrdU in *Gfi-1*^{-/-} (-/-) mice compared with wild-type littermates (+/+). Left plots show gating strategy (representative animals); right plots show BrdU stain (blue) compared with isotype control (red) (averages and s.d. of analysis in five mice per group are given). Asterisks indicate percentage of bone marrow. **b**, Analysis of DNA content in thymocytes, sorted bone marrow myeloid cells and double-sorted HSCs from *Gfi-1*^{-/-} (-/-) mice or wild-type littermates (+/+). The asterisk refers to pools of HSCs sorted from groups of 6–8 mice at 4–6 weeks old. Representative plots, averages and s.d. are shown (statistics: Student's *t*-test). **c**, HSCs (Lin⁻, Sca1⁺, c-kit⁺, Flt3⁻) from *Gfi-1*^{-/-} mice have decreased p21 expression. RNA was isolated from double-sorted HSCs, reverse-transcribed and analysed by real-time PCR. Relative quantity was calculated for each gene from a standard curve generated from bone marrow cDNA and normalized for amount of β-actin transcript. Gene expression in *Gfi-1*^{-/-} HSCs (-/-) and relative littermate control HSCs (+/+) is shown (left panel). Results represent the mean ± s.d. of three amplification reactions. Amplification profiles of representative samples are shown, depicting equal amounts of β-actin (centre) and differential amounts of p21 (right). βA, β-actin; Myc, c-Myc; G2, Gata-2; RFU, relative fluorescence units.

whereas contribution of *Gfi-1*^{-/-} cell DNA was preserved in non-haematopoietic tissues (Fig. 3e). *Gfi-1*^{+/-} ES cells consistently contributed to all tissues (Fig. 3e). Haemoglobin isoform analysis confirmed that progeny of *Gfi-1*^{+/-} but not *Gfi-1*^{-/-} ES cells contributed to haemoglobin production in adult chimaeras supporting an early haematopoietic defect (Fig. 3f). A defect at the level of HSCs was corroborated by the finding that only *Gfi-1*^{+/-} but not *Gfi-1*^{-/-} cells contributed to HPPCs derived from the bone marrow of chimaeric mice (Fig. 3g). Systematic analysis of peripheral blood from chimaeras of different ages revealed that *Gfi-1*^{-/-} ES-cell-derived haemoglobin was detected during the first month of life, but was subsequently lost in all animals (Fig. 3h). Thus, *Gfi-1*^{-/-} HSCs initiate, but do not sustain, bone marrow haematopoiesis in chimaeric mice. Taken together, the transplantation and chimaera experiments provide persuasive evidence that the defective function of *Gfi-1*^{-/-} HSCs is intrinsic to the HSCs themselves rather than a secondary consequence of an impaired haematopoietic microenvironment.

In principle, failure to maintain HSC function, as observed after *Gfi-1* loss in both chimaera and transplant settings, might reflect decreased proliferative capacity and depletion of HSCs, as reported for loss of Bmi-1 (ref. 4). However, the number of phenotypic HSCs in young *Gfi-1*^{-/-} mice was not severely reduced but instead was slightly elevated (Fig. 1a), a finding most consistent with impaired function rather than mere absence. Of note, loss of HSC function has been ascribed to excessive proliferation. Specifically, forced stem cell expansion driven by serial transplantation^{18,19} or by genetic disruption of a negative regulatory component of the cell-cycle machinery²⁰ results in functional compromise without depletion of immunophenotyped HSCs^{19,20}. To address a role for *Gfi-1* in HSC proliferation directly we measured DNA synthesis by determining the fraction of cells that had incorporated 5-bromodeoxyuridine (BrdU) during 3-day *in vivo* exposure (Fig. 4a). Bone marrow was stained with antibodies to identify HSCs (Lin⁻, Sca1⁺, c-kit⁺) and detect nuclear BrdU simultaneously. In agreement with previous data³ ~28% of HSCs from wild-type mice incorporated BrdU and therefore had proliferated during the exposure. Surprisingly, more than double the percentage (67%) of HSCs from *Gfi-1*^{-/-} mice proliferated during the same interval. Furthermore, we demonstrated that HSC hyperproliferation is intrinsic to the haematopoietic system via bone marrow transplantation of *Gfi-1*^{-/-} or control marrow into wild-type mice. As previously reported²¹, we observed that bone marrow transplant of wild-type HSCs resulted in a sustained increase in HSC proliferation, but in the absence of *Gfi-1* this was markedly more pronounced (Supplementary Fig. 2). Interestingly, HSC frequencies were markedly decreased in recipients of *Gfi-1*^{-/-} marrow, potentially due to exhaustion after the added proliferative stress of transplantation (Supplementary Fig. 2b). As expression of *Gfi-1* promotes proliferation of lymphoid cells^{7–9}, we considered whether our results might reflect differences in experimental approach (overexpression versus loss of function) or rather signify authentic context-specific, divergent involvement of *Gfi-1* in controlling cellular proliferation. Therefore, we isolated representative haematopoietic cells from *Gfi-1*^{-/-} and control mice, and assessed cell-cycle stages by measuring DNA content after propidium iodide staining (Fig. 4b). In the thymus of *Gfi-1*^{-/-} mice, in agreement with markedly reduced overall cellularity¹³, the proportion of cells in non-proliferative stages of the cell cycle (G0 and G1) was slightly increased whereas cells in active proliferation (S, G2, M) were reduced in number. These findings are consistent with the previous conclusion that *Gfi-1* promotes proliferation of lymphoid cells. We also analysed Gr1⁺ Mac1⁺ myeloid cells from bone marrow. Granulocytic precursors of *Gfi-1*^{-/-} mice, although arrested in differentiation and not overtly malignant, progressively expand in bone marrow as mice age¹³. Remarkably, these cells in *Gfi-1*^{-/-} mice are highly proliferative as compared with Gr1⁺ Mac1⁺ of wild-type mice. Consistent with the BrdU incorporation

data, the proportion of cycling long-term-repopulating HSCs (Lin⁻, Sca1⁺, c-kit⁺, Flt3⁻)—which were sorted by fluorescence-activated cell sorting (FACS) twice to exclude contamination by other cell types—was appreciably increased in *Gfi-1*^{-/-} mice as compared with controls. Thus, Gfi-1 differentially affects proliferation and cell-cycle progression of haematopoietic cells, dependent on cell lineage and state of differentiation.

To elucidate potential mechanisms through which Gfi-1 might control HSC growth we used quantitative, real-time polymerase chain reaction with reverse transcription (RT-PCR) to assess transcript levels for candidate molecules previously implicated in HSC or progenitor proliferation. Bmi-1, c-Myc and p27 were expressed at similar levels in wild-type and mutant HSCs. Gata-2, which is thought to promote proliferation, was slightly upregulated. Of note, expression of cyclin-dependent kinase inhibitor and G1 checkpoint regulator p21^{Cip1/Waf1} was markedly downregulated (at least tenfold) in *Gfi-1*^{-/-} HSCs (Fig. 4c). Interestingly, p21^{Cip1/Waf1}, a generic regulator of the cell cycle, is required to maintain HSCs (lineage⁻, Hoechst 33342^{low}) in G0 (distinguished from G1 by low RNA content)²⁰. In the absence of p21^{Cip1/Waf1}, HSCs exhibit impaired serial transplantation capacity. Although the effects of p21^{Cip1/Waf1} loss on HSC function are less pronounced than those described here, our data argue that p21^{Cip1/Waf1} is one mediator of Gfi-1 in stem cells. In view of the more extreme phenotype on its loss, we anticipate that Gfi-1 functions through additional pathways.

In keeping with our findings, retroviral insertions into the *Gfi-1* locus (which typically *trans*-activate and cause overexpression) are almost exclusively found in lymphoid tumours in mice¹⁰. Thus far, there is no evidence that loss of Gfi-1 function may also contribute to oncogenesis in other settings by releasing proliferative control at the level of HSCs or after commitment to the myeloid lineage. Our data, however, warrant consideration of this possibility. The versatile role of Gfi-1 in proliferation highlights the critical importance of context in transcriptional regulation *in vivo*, where its function may be modified by cell-type-specific events such as post-translational modifications, interaction with lineage-specific partner molecules, or differential access to regulatory elements of critical target genes. The paralogue of Gfi-1, Gfi-1b, is co-expressed with Gfi-1 in HSCs^{12,22} and is slightly (~twofold) upregulated in *Gfi-1*^{-/-} HSCs (assessed by quantitative PCR; data not shown). Yet, we demonstrated by competitive repopulation and serial transplantation assays using donor cells from *Gfi-1b*^{-/-} fetal liver that in marked contrast to Gfi-1, Gfi-1b is dispensable for HSC function (Supplementary Fig. 3).

The severe disturbance of HSC function that we have documented in the absence of Gfi-1 is probably the result of excessive cycling. Yet, we cannot exclude the possibility that it reflects additional roles of Gfi-1 in HSCs. Nevertheless, our data provide molecular insight into how the transcriptional machinery protects HSCs by curbing excessive or unnecessary proliferation. Elucidation of the role of Gfi-1 in HSCs provides a novel entry point into the molecular mechanics of the HSC 'brake' for proliferation, and may ultimately generate opportunities for its manipulation. □

Methods

Flow cytometry

MoFlo, FACS-Calibur or FACS-Scan flow cytometers were used for analysis and sorting. Antibody conjugates and matched isotype controls were obtained from PharMingen or eBioscience. HSC stain: lineage markers CD4 (RM4-5), CD8 (53-6), CD3e (145-2C11), TCRβ (H54-597), B220 (RA3-6B2), CD19 (6D5), Gr1 (RB6-8C5), Mac-1 (M1/70), Ter119 (all Cy-Chrome or PE-Cy5); CD117 (c-kit, 2B8, APC), CD135 (Flt3, A2F10, PE), Sca-1 (E13-161, fluorescein isothiocyanate), propidium iodide (1 µg ml⁻¹; Molecular Probes). CD45 isotype analysis: B220 (RA3-6B2, APC), Gr1 (RB6-8C5, APC), Mac1 (M1/70, PE), CD45.1 (104, fluorescein isothiocyanate), CD45.2 (A20, PE, fluorescein isothiocyanate).

Bone marrow transplantation

Bone marrow transplantations were performed by intravenous injection of donor cells from 4-week-old *Gfi-1*^{-/-} mice or littermates (C57bl6/129 mixed) and/or competitor

marrow (B6.SJL; 002014, Jackson) into recipients that had received 1,200 cGy of radiation as a split dose. Donor cells from *Gfi-1b*^{-/-} fetal livers were obtained on day 12.5 of embryonic development after timed mating of heterozygotes²². Recipients were B6/129SF1/J (101043; Jackson) in experiments shown in Fig. 2. Similar results were obtained using B6.SJL mice as recipients, which were also used in experiments shown in Supplementary Figs 1, 2 and 3. Southern blot, automated blood counts and bone marrow morphology was as described¹³.

Colony-forming assays

HPPCs were generated by seeding bone marrow cells into methylcellulose-containing media (MethoCult 3234; Stem Cell Technologies) in the presence of SCF, interleukin (IL)-3, EPO, IL-6, IL-1α granulocyte-macrophage colony-stimulating factor and macrophage colony-stimulating factor (all R&D Systems). Colonies with a tight centre and a diameter ≥ 1 mm at day 12 were isolated using a capillary under an inverted microscope, pooled and processed for FACS staining or DNA extraction.

Generation of chimaeras

Targeting of Gfi-1 in ES cells has been described¹³. ES cells homozygous for Gfi-1 deletion were obtained by selecting clones that grew despite increased neomycin concentration (1.7 mg ml⁻¹) owing to crossing over of the targeted allele. Subsequent Southern blot analysis confirmed duplication of the targeted mutation and absence of wild-type Gfi-1 (probe B¹³). Chimaeras were obtained by injecting *Gfi-1*^{+/+} or *Gfi-1*^{-/-} ES clones into C57/bl6 blastocysts. Chimaerism was assessed by agouti coat-colour contribution. Only animals with contribution >70% were selected for analysis. Southern blot analysis of selected tissues after DNA extraction (internal probe¹³) and haemoglobin isoform analysis was performed as described²³.

BrdU incorporation

Analysis of BrdU incorporation was performed using the FITC BrdU Flow Kit (PharMingen) after a single intraperitoneal injection of BrdU (PharMingen, 1 mg per 6 g of mouse weight) and admixture of 1 mg ml⁻¹ of BrdU (Sigma, 1 mg ml⁻¹) to drinking water for 3 days (or 2 days after transplant). Mice were 4–6 weeks of age. Surface stain consisted of lineage markers as above, CD117 (c-kit, 2B8, APC) and Sca-1 (E13-161, PE).

Cell-cycle analysis

Cell-cycle analysis of thymocytes, or sorted myeloid and HSC populations was performed after ethanol fixation, RNase digestion and propidium iodide staining (Roche, 1399861) using BD ModFit LT software.

Gene expression analysis

RNA was isolated from HSCs (8,100 cells; derived from pooled bone marrow of 6–7 mice) using the RNeasy Micro Kit (Qiagen) and recommended protocol. First-strand complementary DNA synthesis was primed with OligodT (Invitrogen) from sample RNA (2 µl) or wild-type bone marrow RNA (100 ng) using the Sensiscript RT Kit (Qiagen). For real-time PCR analysis, cDNA (1.5 µl) was mixed with iQ SYBR green supermix reagent (BioRad), primers (Supplementary Table 1) and amplification plots were generated using the iCycler Thermal cycler and iQ 4-colour real-time detection system (BioRad). To generate standard curves, cDNA from wild-type bone marrow RNA (100 ng) was used as template in a threefold dilution series (0–81-fold). Sample cDNA was used undiluted. Relative expression was calculated using the standard curve method (ABI Prism 7700 Sequence Detection System, user Bulletin 2, PE Applied Biosystems). Primers as in Supplementary Table 1.

Received 14 May; accepted 10 August 2004; doi:10.1038/nature02994.

Published online 29 September 2004.

- Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
- Jordan, C. T. & Lemischka, I. R. Clonal and systemic analysis of long-term hematopoiesis in the mouse. *Genes Dev.* **4**, 220–232 (1990).
- Cheshier, S. H., Morrison, S. J., Liao, X. & Weissman, I. L. *In vivo* proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **96**, 3120–3125 (1999).
- Park, I. K. *et al.* Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302–305 (2003).
- Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255–260 (2003).
- Sauvageau, G. *et al.* Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations *in vitro* and *in vivo*. *Genes Dev.* **9**, 1753–1765 (1995).
- van Lohuizen, M. *et al.* Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* **65**, 737–752 (1991).
- Gilks, C. B., Bear, S. E., Grimes, H. L. & Tschlis, P. N. Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma lines to IL-2-independent growth following activation of a gene (Gfi-1) encoding a novel zinc finger protein. *Mol. Cell. Biol.* **13**, 1759–1768 (1993).
- Zhu, J. *et al.* Growth factor independent-1 induced by IL-4 regulates Th2 cell proliferation. *Immunity* **16**, 733–744 (2002).
- Akagi, K., Suzuki, T., Stephens, R. M., Jenkins, N. A. & Copeland, N. G. RTGCD: retroviral tagged cancer gene database. *Nucleic Acids Res.* **32**, D523–D527 (2004).
- Scheijen, B., Jonkers, J., Acton, D. & Berns, A. Characterization of pal-1, a common proviral insertion site in murine leukemia virus-induced lymphomas of c-myc and Pim-1 transgenic mice. *J. Virol.* **71**, 9–16 (1997).
- Phillips, R. L. *et al.* The genetic program of hematopoietic stem cells. *Science* **288**, 1635–1640 (2000).
- Hock, H. *et al.* Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. *Immunity* **18**, 109–120 (2003).

14. Chang, H., Jensen, L. A., Quesenberry, P. & Bertoncello, I. Standardization of hematopoietic stem cell assays: a summary of a workshop and working group meeting sponsored by the National Heart, Lung, and Blood Institute held at the National Institutes of Health, Bethesda, MD on September 8–9, 1998 and July 30, 1999. *Exp. Hematol.* **28**, 743–752 (2000).
15. Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* **98**, 14541–14546 (2001).
16. Adolfsson, J. *et al.* Upregulation of Flt3 expression within the bone marrow Lin(–)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* **15**, 659–669 (2001).
17. Yucel, R., Karsunky, H., Klein-Hitpass, L. & Moroy, T. The transcriptional repressor Gfi1 affects development of early, uncommitted c-Kit + T cell progenitors and CD4/CD8 lineage decision in the thymus. *J. Exp. Med.* **197**, 831–844 (2003).
18. Ross, E. A., Anderson, N. & Micklem, H. S. Serial depletion and regeneration of the murine hematopoietic system. Implications for hematopoietic organization and the study of cellular aging. *J. Exp. Med.* **155**, 432–444 (1982).
19. Spangrude, G. J., Brooks, D. M. & Tumas, D. B. Long-term repopulation of irradiated mice with limiting numbers of purified hematopoietic stem cells: *in vivo* expansion of stem cell phenotype but not function. *Blood* **85**, 1006–1016 (1995).
20. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science* **287**, 1804–1808 (2000).
21. Allsopp, R. C., Cheshier, S. & Weissman, I. L. Telomere shortening accompanies increased cell cycle activity during serial transplantation of hematopoietic stem cells. *J. Exp. Med.* **193**, 917–924 (2001).
22. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev.* **16**, 301–306 (2002).
23. Tsai, F. Y. *et al.* An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221–226 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We appreciate the assistance of M. Handley, H. Levine, J. LaVechio, C. Browne and A. Williams. H.H. was supported by an NCI Career Development award. S.H.O. is an Investigator of the Howard Hughes Medical Institute. This work was supported in part by a Center of Excellence in Molecular Hematology award from the NIH-NIDDK.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H.O. (stuart_orkin@dfci.harvard.edu).

Essential role for the p110 δ phosphoinositide 3-kinase in the allergic response

Khaled Ali^{1,2}, Antonio Bilancio¹, Matthew Thomas³, Wayne Pearce¹, Alasdair M. Gilfillan⁴, Christine Tkaczyk⁴, Nicolas Kuehn⁵, Alexander Gray⁶, June Giddings³, Emma Peskett¹, Roy Fox³, Ian Bruce³, Christoph Walker³, Carol Sawyer^{1*}, Klaus Okkenhaug^{1*}, Peter Finan³ & Bart Vanhaesebroeck^{1,2}

¹Ludwig Institute for Cancer Research, 91 Riding House Street, London W1W 7BS, UK

²Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

³Novartis Institutes for BioMedical Research, Horsham RH12 5AB, UK

⁴Laboratory of Allergic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Building 10, Room 11C206, 10 Center Drive, MSC 1881, Bethesda, Maryland 20892-1881, USA

⁵Frimorfo, PO BOX 191, CH-1705 Fribourg, Switzerland

⁶Department of Biochemistry, University of Dundee, Dundee DD1 5EH, UK

* Present addresses: Cancer Research UK Centre for Cancer Therapeutics, Institute of Cancer Research, 15 Cotswold Road, Belmont, Sutton SM2 5NG, UK (C.S.); the Babraham Institute, Cambridge, CB2 4AT, UK (K.O.)

Inflammatory substances released by mast cells induce and maintain the allergic response^{1,2}. Mast cell differentiation and activation are regulated, respectively, by stem cell factor (SCF; also known as Kit ligand) and by allergen in complex with allergen-specific immunoglobulin E (IgE)^{2,3}. Activated SCF receptors and high-affinity receptors for IgE (Fc ϵ RI) engage phosphoinositide 3-kinases (PI(3)Ks) to generate intracellular

lipid second messenger signals^{2–5}. Here, we report that genetic or pharmacological inactivation of the p110 δ isoform of PI(3)K in mast cells leads to defective SCF-mediated *in vitro* proliferation, adhesion and migration, and to impaired allergen–IgE-induced degranulation and cytokine release. Inactivation of p110 δ protects mice against anaphylactic allergic responses. These results identify p110 δ as a new target for therapeutic intervention in allergy and mast-cell-related pathologies.

PI(3)Ks activated by the Fc ϵ RI and SCF receptors include the heterodimeric class IA PI(3)Ks, which consist of a p110 α , p110 β or p110 δ catalytic subunit bound to either one of five regulatory proteins (p85 α , p55 α , p50 α , p85 β or p55 γ) that link the p110 subunits to tyrosine kinase signalling pathways⁶. Mast cells express high levels of the p110 δ isoform⁷ which, unlike the ubiquitously expressed p110 α and p110 β , is primarily found in leukocytes^{8,9}. Whereas mice lacking p110 α or p110 β are embryonic lethal⁶, p110 δ -null mice are viable, with specific defects in B and T cells¹⁰. To investigate the role of p110 δ in mast cells, we derived bone marrow mast cells (BMMCs) from mice expressing either wild-type p110 δ or p110 δ ^{D910A}, a loss-of-function allele of p110 δ (ref. 11). Culture of bone marrow precursor cells from both genotypes in the presence of interleukin-3 (IL-3) gave rise to mature BMMCs, which expressed identical levels of Fc ϵ RI and Kit receptor (as assessed by fluorescence-activated cell sorting analysis; data not shown) and were morphologically indistinguishable (as assessed by alcian blue and safranin staining; data not shown). However, the number of BMMCs obtained from p110 δ ^{D910A/D910A} cultures was substantially lower (up to 60%; data not shown). Similar observations were made when precursor cells were cultured in the presence of IL-3 plus SCF, giving rise to up to 50% less BMMCs in p110 δ mutant cells (data not shown).

The level of class I PI(3)K isoform expression in p110 δ ^{D910A/D910A} BMMCs is similar to that of wild-type cells (Fig. 1a). p110 δ lipid

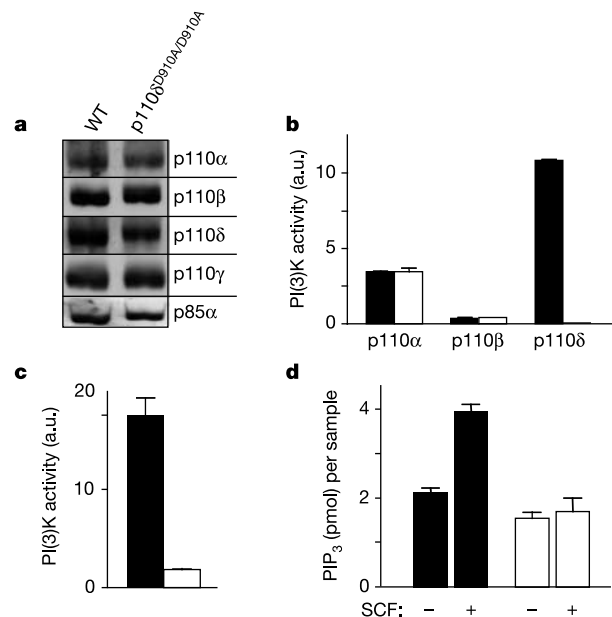


Figure 1 Effect of p110 δ inactivation on PI(3)K expression and activity. **a**, Expression of class I PI(3)K isoforms in total cell lysates (50 μ g per lane), assessed by immunoblotting. WT, wild type. **b**, *In vitro* lipid kinase activity of class IA PI(3)Ks. p110 isoforms were immunoprecipitated using isoform-specific antibodies, and tested for associated lipid kinase activity (presented as arbitrary units (a.u.)). **c**, *In vitro* lipid kinase activity of BMMC class IA PI(3)Ks, isolated using Y_pVPMLG (where Y_p is phosphotyrosine) peptide complexes. **d**, *In vivo* PIP₃ levels in BMMCs stimulated with or without 100 ng ml⁻¹ SCF for 10 min. Filled bars, wild type; open bars, p110 δ ^{D910A/D910A}.

kinase activity was abrogated in p110 $\delta^{D910A/D910A}$ BMMCs, with no alterations in the kinase activities of p110 α and p110 β (Fig. 1b). Total *in vitro* class IA PI(3)K lipid kinase activity, isolated from mutant BMMCs, was reduced by up to 90% (Fig. 1c), indicating that p110 δ contributes substantially to the overall class IA PI(3)K activity in mast cells. In agreement with this, there was no increase in *in vivo* phosphatidylinositol 3,4,5-trisphosphate (PIP₃) above resting levels in p110 $\delta^{D910A/D910A}$ cells upon stimulation with SCF (Fig. 1d).

To assess the impact of p110 δ inactivation on intracellular signalling, we monitored Ser 473 phosphorylation of protein kinase B (PKB, also known as Akt) and Tyr 204 phosphorylation of Erk as indirect measures of activation of these kinases. SCF-induced PKB phosphorylation was almost completely abrogated in p110 $\delta^{D910A/D910A}$ BMMCs, in contrast with Erk activation, which was not affected (Fig. 2a). These results could be mimicked in wild-type cells by the p110 δ -selective inhibitor IC87114 (ref. 12) (Fig. 2c, d). It has been reported¹² that IC87114 is selective for p110 δ in cells at doses up to 5–10 μ M; this is in line with our observation that such doses of this compound do not affect cells in which p110 δ is already genetically inactivated (data not shown and Figs 2f and 3c). IL-3-induced PKB phosphorylation was severely reduced but not completely abrogated in mutant mast cells (Fig. 2b). IL-3-stimulated PKB phosphorylation could be completely blocked by treatment with LY294002, a broad-spectrum PI(3)K inhibitor

(Fig. 2e and ref. 13), indicating that isoforms other than p110 δ contribute to PI(3)K signalling downstream of the IL-3 receptor. The residual IL-3-induced PI(3)K activity in p110 $\delta^{D910A/D910A}$ cells may explain why mature BMMCs can be derived from p110 $\delta^{D910A/D910A}$ bone marrow precursors in the presence of this cytokine. As was seen for SCF, IL-3-induced Erk activation was not affected by inactivation of p110 δ (Fig. 2b), and pharmacological inhibition of p110 δ using IC87114 on wild-type cells mirrored the impact of genetic inactivation of p110 δ on IL-3 signalling (Fig. 2e).

We next assessed *in vitro* functional responses of p110 $\delta^{D910A/D910A}$ mast cells to SCF or IL-3. DNA synthesis induced by each cytokine was severely reduced (Fig. 2f, left and middle panels). The synergistic effect on DNA synthesis of co-stimulation with IL-3 and SCF was also substantially diminished in p110 $\delta^{D910A/D910A}$ BMMCs (Fig. 2f, right panel). Treatment with IC87114 reduced DNA synthesis of wild-type BMMCs without significantly affecting p110 $\delta^{D910A/D910A}$ BMMCs (Fig. 2f). Other SCF-stimulated responses that were severely decreased in p110 $\delta^{D910A/D910A}$ mast cells include the ability to adhere to fibronectin upon SCF stimulation (Fig. 2g) and migration towards SCF (Fig. 2h). Taken together, these data indicate a critical contribution of p110 δ in signalling of SCF towards physiological functions of this cytokine such as proliferation, stimulation of adhesion to the extracellular matrix and chemotactic activity for homing of mast cells to tissues¹⁴.

We next compared the capacity of BMMCs from p110 $\delta^{D910A/D910A}$

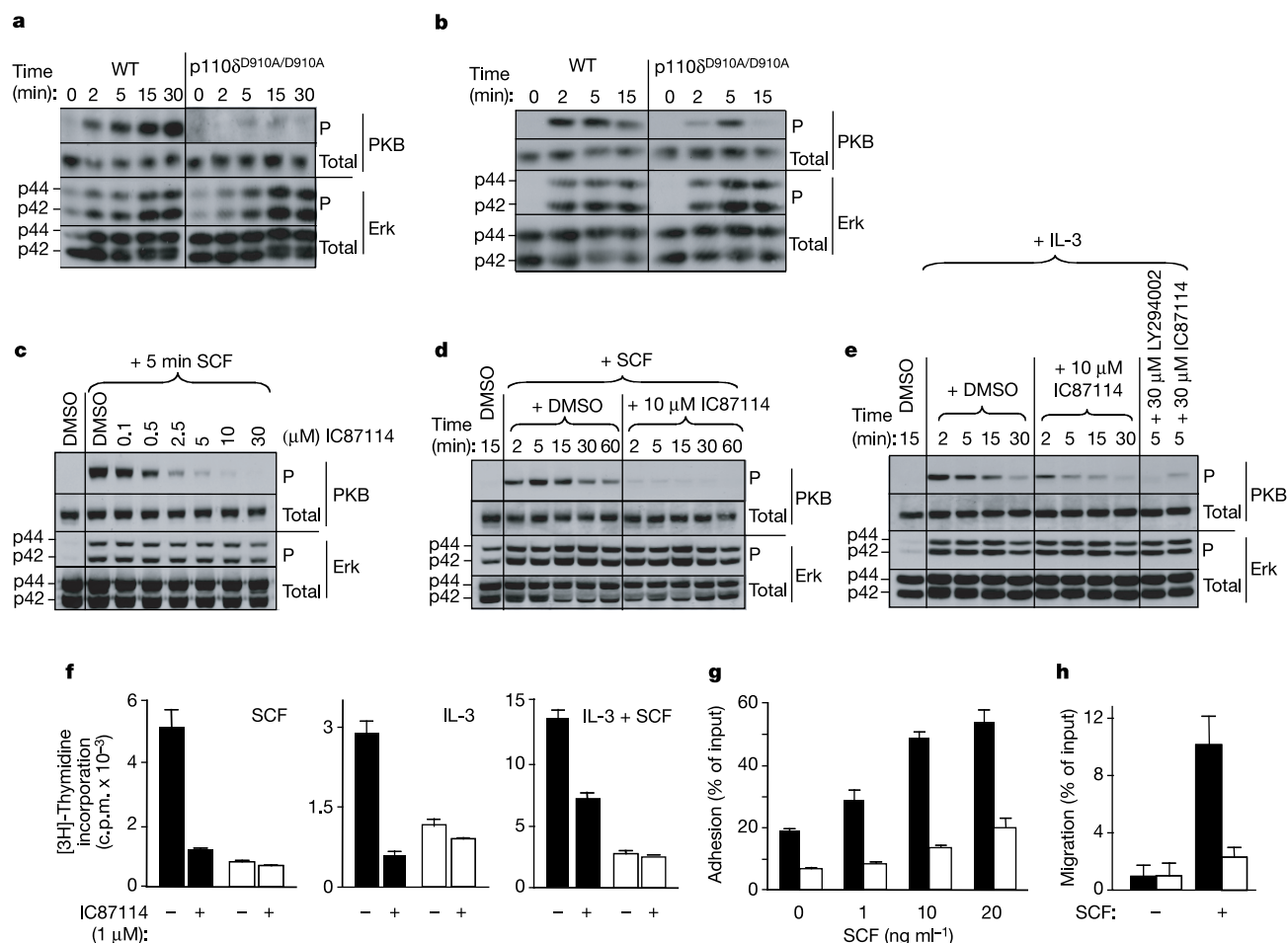


Figure 2 Effect of p110 δ inactivation on SCF- or IL-3-induced BMMC responses.

a, Signalling in BMMCs stimulated with SCF. **b**, Signalling in BMMCs stimulated with IL-3. **c**, **d**, Effect of IC87114 on SCF signalling in wild-type BMMCs. **e**, Effect of IC87114 and LY294002 on IL-3 signalling in wild-type BMMCs. Inhibitors were added 20 min before

stimulation. **f**, Effect of p110 δ inactivation on SCF- and IL-3-induced DNA synthesis after 48 h. **g**, **h**, Effect of p110 δ inactivation on SCF-induced adhesion and migration. Filled bars, wild type; open bars, p110 $\delta^{D910A/D910A}$. Data in **f–h** are expressed as mean $\pm$ s.d.

and wild-type mice to respond to antigen in complex with IgE. Such complexes aggregate the Fc ϵ RI on mast cells, initiating biochemical signalling cascades that ultimately lead to the release of inflammatory mediators, cytokines and to much of the pathology associated with the allergic response^{1,2,4}. Ser 473 phosphorylation of PKB induced by Fc ϵ RI activation was substantially reduced in p110 $\delta^{D910A/D910A}$ BMMCs (Fig. 3a). We also tested the capacity of cultured mast cells to release contents from intracellular stores into the surrounding environment upon stimulation with antigen/allergen via the Fc ϵ RI receptor. As a read-out, we measured release of β -hexosaminidase, an enzyme stored in mast cell granules¹⁵. Inclusion of 30 μ M of LY294002, a broad-spectrum PI(3)K inhibitor, reduced the degranulation of wild-type mast cells by ~80% (Fig. 3b, left panel), confirming that PI(3)K is a significant component of the signalling cascade leading to mast cell degranulation^{15,16}. p110 $\delta^{D910A/D910A}$ BMMCs showed a 45–55% reduction in Fc ϵ RI-induced degranulation, relative to wild-type cells (Fig. 3b, left panel). Similar results were obtained when the cells were incubated with IgE overnight, washed and then stimulated with antigen (Ag), with degranulation defects observed across a wide dose range of antigen (Fig. 3b, right panel). Degranulation in p110 $\delta^{D910A/D910A}$ BMMCs could be further reduced by LY294002, as in wild-type cells (left panel of Fig. 3b). Pre-incubation of wild-type mast cells with 5 μ M IC87114 reduced degranulation levels to those seen in p110 $\delta^{D910A/D910A}$ cells while having no effect on p110 δ mutant cells (Fig. 3c). IC87114 inhibited degranulation to the same extent in human mast cells (Fig. 3d) and also blocked the capacity of SCF to potentiate degranulation by low doses of antigen (Fig. 3e). The observation that p110 δ is critical for the SCF component of mast cell degranulation is of particular physiological relevance, given that mast cells are continuously exposed to SCF in the local tissue environment. Mast cell activation thus occurs in the context of SCF signalling, especially during inflammatory response with excessive local production of SCF^{17,18}. p110 $\delta^{D910A/D910A}$ BMMCs also showed defects in secretion of inflammatory cytokines such as

TNF- α and IL-6 (Fig. 3f). Overall, these data indicate that p110 δ is the main component of the PI(3)K-dependent antigen–IgE signalling cascade leading to degranulation, and is essential for the SCF-mediated potentiation of Fc ϵ RI activity.

In p110 $\delta^{D910A/D910A}$ mice mast cell numbers were differentially affected in distinct anatomical locations (Supplementary Table 1), with defects ranging from severe (peritoneum, jejunum, ileum and colon) to moderate (ear dermis (Fig. 4a), stomach submucosa and muscularis) to no significant differences (back dermis (Fig. 4a), mucosa of the stomach). A decrease in mast cell numbers is in line with a critical role of SCF in mast cell differentiation and survival, and the importance of p110 δ in signalling by this cytokine. The residual mast cells present in p110 $\delta^{D910A/D910A}$ mice may result from SCF signalling pathways unaffected by p110 δ inactivation (such as the Erk pathway) and/or from developmental/homeostatic cues other than SCF.

We next investigated the impact of p110 δ inactivation on allergic responses in the mouse. *In vivo* anaphylaxis through the Fc ϵ RI receptor is an allergic response that is predominantly mast-cell-dependent and occurs as a result of either local or systemic exposure to allergens, which crosslink and activate antigen-specific IgE bound to the Fc ϵ RI on the mast cell surface. We assessed the capacity of wild-type and p110 $\delta^{D910A/D910A}$ mice to respond to immune challenge in a passive cutaneous anaphylaxis (PCA) model¹⁹. Mice were given an intradermal injection of IgE directed against a hapten (dinitrophenyl (DNP) in this case). Between 24–48 h after this priming event the mice were challenged by systemic administration of DNP coupled to a carrier protein (HSA) together with Evan's blue dye, which binds to serum proteins. Mediators released upon mast cell activation increase vascular permeability, which allows the dye to leak from the blood vessels into the surrounding tissue, causing oedema and providing a measure of mast cell activation. We assessed the PCA response in two distinct tissue locations. In the ear of p110 $\delta^{D910A/D910A}$ mice, a marked (80%) reduction in PCA was observed (Fig. 4b). This decrease is more significant than the

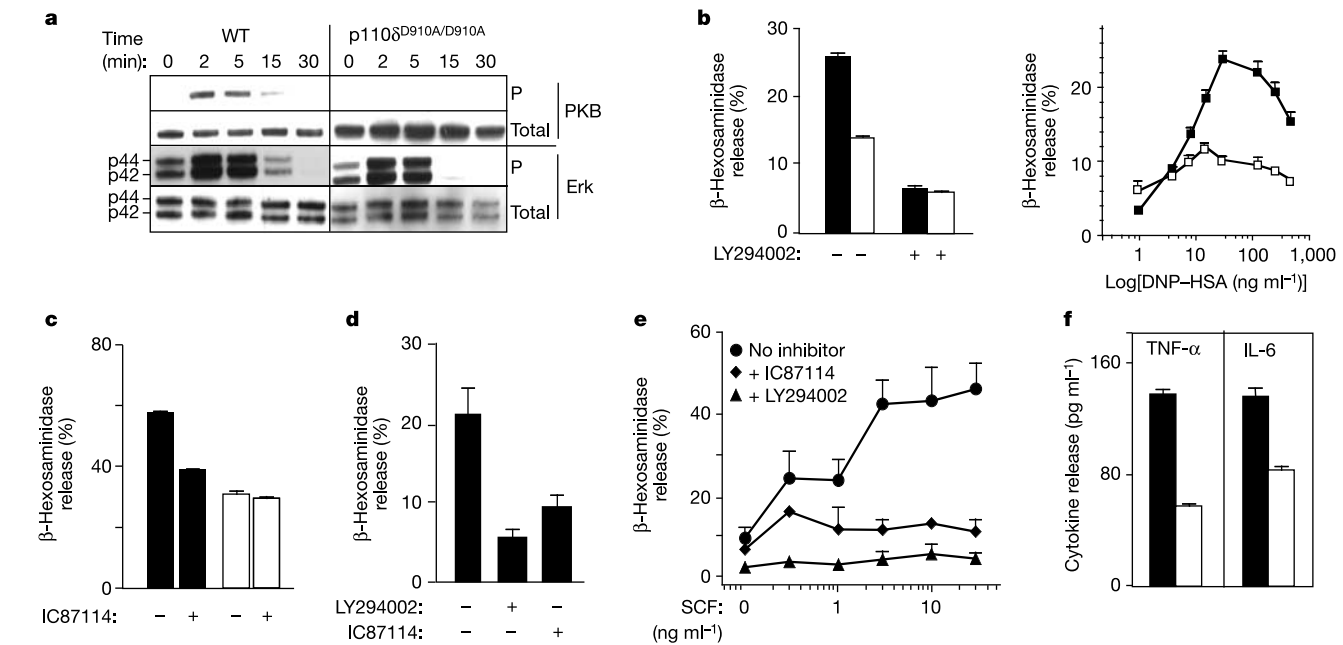


Figure 3 Effect of p110 δ inactivation on Fc ϵ RI-induced BMMC responses. **a–c**, Signalling (**a**) and β -hexosaminidase release (**b**, **c**) in BMMCs stimulated with IgE–antigen. Right panel of **b**: dose titration of DNP–HSA. **d**, Effect of inhibitors on degranulation of human mast cells challenged with antigen (1 ng ml⁻¹). **e**, Effect of PI(3)K inhibitors on SCF-potentiated human mast cell degranulation (using a suboptimal antigen

dose of 0.1 ng ml⁻¹). **f**, Cytokine release from BMMCs stimulated with IgE–antigen. LY294002 and IC87114 were used at 30 and 5 μ M, respectively, and added 20 min before IgE–antigen stimulation. Filled bars, wild type; open bars, p110 $\delta^{D910A/D910A}$. Data in **b–f** are expressed as mean $\pm$ s.d.

reduction in mast cell numbers at this location (maximally 47%; Supplementary Table 1), suggesting that the remaining mutant mast cells may be defective. PCA in the back dermis—a site with unaltered mast cell numbers in p110 $\delta^{D910A/D910A}$ mice (Supplementary Table 1; Fig. 4a)—was also significantly reduced (40%; Fig. 4c). IC87114, at pharmacological doses at which it is selective for p110 δ (plasma concentration of 5 μ M from a 15 mg kg $^{-1}$ dose), reduced the allergic response by 35–40% in the back skin and ear of wild-type mice (Fig. 4d; data for the ear are not shown). At higher plasma concentrations (40 μ M from a 60 mg kg $^{-1}$ dose) IC87114 completely blocked the anaphylaxis reaction in the back dermis (Fig. 4d).

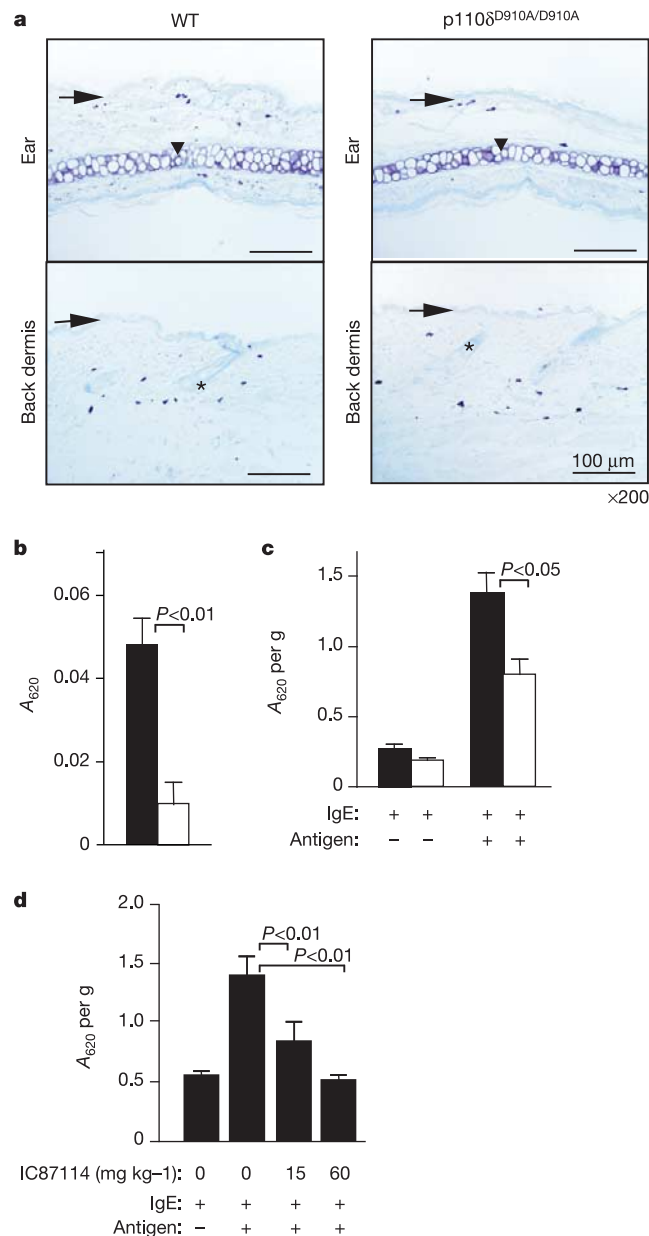


Figure 4 Effect of p110 δ inactivation *in vivo*. **a**, Reduced numbers of mast cells (stained with toluidine blue) in the ear dermis but not in the back skin. Arrows, epidermis; arrowheads, cartilage; asterisk, hair follicles with sebaceous glands. **b**, PCA response (ear) of wild-type ($n = 13$) and p110 $\delta^{D910A/D910A}$ mice ($n = 10$). **c**, PCA response (back skin) of mice injected with anti-DNP IgE, challenged with saline ($n = 4$ for wild type; $n = 5$ for p110 $\delta^{D910A/D910A}$) or DNP-HSA ($n = 5$ for wild type; $n = 6$ for p110 $\delta^{D910A/D910A}$). **d**, IC87114 attenuates the PCA response (back skin) of wild-type mice ($n = 8$ for each treatment group). Filled bars, wild type; open bars, p110 $\delta^{D910A/D910A}$.

This may be a consequence of the inhibition of other PI(3)K isoforms (such as p110 γ), which could account for the remaining PI(3)K-dependent (that is, LY294002-sensitive) degranulation seen in p110 $\delta^{D910A/D910A}$ BMMCs (Fig. 3b, left panel). These data show that this class of small molecule inhibitors of PI(3)K can completely abrogate *in vivo* cutaneous anaphylaxis.

Taken together, our data show that the p110 δ PI(3)K isoform has an important role in mast cell homeostasis and in the allergic response. Our observations are in contrast with studies in mice lacking regulatory subunits of class IA PI(3)Ks, which indicated that class IA PI(3)K enzymes are not involved in IL-3 and Fc ϵ RI-mediated allergic responses^{7,15,16,20}. The reason for this discrepancy is not clear, but may relate to the complex impact of loss of expression of class IA PI(3)K regulatory subunits on the class IA PI(3)K system, with increased expression of the remaining regulatory subunits and reduced expression of all p110 isoforms, and even increased PI(3)K signalling under certain circumstances^{21–25}. This contrasts with the present study, in which p110 δ was selectively inactivated by introduction of a point mutation in the ATP-binding site, mimicking the effect of a systemically administered ATP-competitive p110 δ inhibitor. Indeed, *in vivo* administration of a p110 δ -selective inhibitor reproduced the therapeutic effect of the D910A mutation in the context of anaphylaxis. Interference with PI(3)K signalling pathways is being explored as a means of therapeutic intervention in many different pathological conditions²⁶. Our work suggests a promising therapeutic potential for p110 δ -selective compounds to treat allergy and other pathological conditions in which mast cells have a central role. □

Methods

Mice

p110 $\delta^{D910A/D910A}$ mice have been described previously¹¹ and were backcrossed on a C57BL/6 or BALB/c background for 6–10 generations. Age-matched, 6–10-week-old littermate mice were used for all experiments.

Mast cell cultures

Total bone marrow samples from the femurs of 3–5 mice were cultured in RPMI medium supplemented with 10% ultra-low IgG FBS (Gibco), 10 ng ml $^{-1}$ IL-3 (Tebu-bio), 20 ng ml $^{-1}$ SCF (Tebu-bio), 2 mM L-glutamine, 100 U ml $^{-1}$ penicillin and streptomycin (Gibco) in a 5% CO $_2$ atmosphere. After a 4-week maturation period, BMMCs were collected and tested for c-kit receptor and Fc ϵ RI expression by flow cytometry using phycoerythrin-labelled rat anti-c-kit (Pharmingen) or anti-DNP IgE (SPE-7; Sigma) and monoclonal fluorescein isothiocyanate-anti-mouse IgE (Molecular Probes) antibodies, respectively. BMMCs used for this study were cultured with IL-3 and SCF, unless otherwise indicated, with culture times not exceeding 6 weeks.

Cytokine stimulation, lysis and immunoblotting

Cells were starved for 24 h in normal culture medium lacking IL-3 and SCF. After stimulation with 20 ng ml $^{-1}$ IL-3 and/or 20 ng ml $^{-1}$ SCF, the cells were washed twice with ice-cold PBS, lysed and processed for immunoblotting and kinase assays as described¹¹.

Lipid kinase assay

In vitro lipid kinase assays were carried out as previously described¹¹ (see Supplementary Information for detailed details). *In vivo* levels of PIP $_3$ were determined by a time-resolved fluorescence resonance energy transfer ligand displacement assay, as described²⁷.

DNA synthesis

After a 24-h starvation period in culture medium without IL-3 and SCF, 10 5 cells in triplicate were incubated with 20 ng ml $^{-1}$ IL-3 and/or 20 ng ml $^{-1}$ SCF, together with [3 H]-thymidine and collected 24–72 h later.

Degranulation and cytokine secretion

Degranulation experiments were carried out in triplicate in 96-well plates. Exponentially growing cells were stimulated with IgE-antigen complex for 60 min. Degranulation is expressed as a percentage of total β -hexosaminidase activity in the input cells (see Supplementary Information for detailed methods). Cytokine released into the supernatants was assessed using Quantikine kits (R&D Systems).

Human mast cell studies

Human mast cells were obtained by culture of peripheral blood CD34 $^+$ cells in Stem Pro 34 media (Invitrogen) containing recombinant human IL-3 (30 ng ml $^{-1}$), IL-6 (100 ng ml $^{-1}$) and SCF (100 ng ml $^{-1}$) (Pepro Tech) for 1 week, followed by culture without IL-3 (but with IL-6 and SCF as above) for a further 6–7 weeks as described²⁸ (see Supplementary Information for detailed methods).

Cell adhesion and migration

For adhesion assays, cells starved of cytokines for 24 h were washed and re-suspended in culture medium without cytokines. Cells were plated at 5×10^4 per well in a 50- μ l volume in 96-well plates pre-coated overnight with $5 \mu\text{g ml}^{-1}$ human plasma fibronectin (Gibco). After 30 min of stimulation with or without SCF (in 100 μ l final volume), non-adherent cells were removed by inverting plates for 15 min. The adherent cells were lysed using HTAB and β -hexosaminidase activity measured to assess the fraction of adherent cells. Adhesion is expressed as percentage of input. For migration assays, exponentially growing cells were washed and re-suspended in Tyrodes buffer. 10^5 cells were seeded in 100 μ l in the upper chamber of 8- μ m-pore-diameter transwells (BD Biosciences) with or without 100 ng ml^{-1} SCF in Tyrodes buffer in the lower chamber. After a 3–4-h incubation at 37°C and 5% CO_2 , the cells in the lower chamber were lysed using HTAB and β -hexosaminidase activity measured to assess the fraction of migrated cells.

Passive cutaneous anaphylaxis

The PCA protocols were adapted from refs 29, 30 (see Supplementary Information for detailed protocol).

Statistical analysis

All *in vitro* data shown are representative experiments (mean $\pm$ s.d.) from different BMMC cultures (established from at least 3–5 littermate mice each). Data for *in vitro* experiments were statistically analysed using a *t*-test and differences between wild-type and p110 $\delta^{\text{D910A/D910A}}$ BMDCs were statistically significant ($P < 0.05$) unless otherwise stated. Results from *in vivo* experiments (mean $\pm$ s.e.m.) were assessed using a Mann–Whitney *U*-test with results of analysis and animal numbers presented in the relevant figure legends. GraphPad Prism software was used for all statistical analyses.

Received 16 July; accepted 2 September 2004; doi:10.1038/nature02991.

- Wedemeyer, J. & Galli, S. J. Mast cells and basophils in acquired immunity. *Br. Med. Bull.* **56**, 936–955 (2000).
- Kinet, J. P. The high-affinity IgE receptor (Fc ϵ RI): from physiology to pathology. *Annu. Rev. Immunol.* **17**, 931–972 (1999).
- Besmer, P. The kit ligand encoded at the murine Steel locus: a pleiotropic growth and differentiation factor. *Curr. Opin. Cell Biol.* **3**, 939–946 (1991).
- Rivera, J. *et al.* Macromolecular protein signaling complexes and mast cell responses: a view of the organization of IgE-dependent mast cell signaling. *Mol. Immunol.* **38**, 1253–1258 (2002).
- Serve, H. *et al.* Differential roles of PI3-kinase and Kit tyrosine 821 in Kit receptor-mediated proliferation, survival and cell adhesion in mast cells. *EMBO J.* **14**, 473–483 (1995).
- Vanhaesebroeck, B. *et al.* Synthesis and function of 3-phosphorylated inositol lipids. *Annu. Rev. Biochem.* **70**, 535–602 (2001).
- Lu-Kuo, J. M., Fruman, D. A., Joyal, D. M., Cantley, L. C. & Katz, H. R. Impaired kit- but not Fc ϵ RI-initiated mast cell activation in the absence of phosphoinositide 3-kinase p85 α gene products. *J. Biol. Chem.* **275**, 6022–6029 (2000).
- Vanhaesebroeck, B. *et al.* p110 δ , a novel phosphoinositide 3-kinase in leukocytes. *Proc. Natl Acad. Sci. USA* **94**, 4330–4335 (1997).
- Chantry, D. *et al.* p110 δ , a novel phosphatidylinositol 3-kinase catalytic subunit that associates with p85 and is expressed predominantly in leukocytes. *J. Biol. Chem.* **272**, 19236–19241 (1997).
- Okkenhaug, K. & Vanhaesebroeck, B. PI3K in lymphocyte development, differentiation and activation. *Nature Rev. Immunol.* **3**, 317–330 (2003).
- Okkenhaug, K. *et al.* Impaired B and T cell antigen receptor signaling in p110 δ PI 3-kinase mutant mice. *Science* **297**, 1031–1034 (2002).
- Sadhu, C., Dick, K., Tino, W. T. & Staunton, D. E. Selective role of PI3K delta in neutrophil inflammatory responses. *Biochem. Biophys. Res. Commun.* **308**, 764–769 (2003).
- Hinton, H. J. & Welham, M. J. Cytokine-induced protein kinase B activation and Bad phosphorylation do not correlate with cell survival of hemopoietic cells. *J. Immunol.* **162**, 7002–7009 (1999).
- Tan, B. L. *et al.* Genetic evidence for convergence of c-Kit- and $\alpha 4$ integrin-mediated signals on class IA PI-3kinase and the Rac pathway in regulating integrin-directed migration in mast cells. *Blood* **101**, 4725–4732 (2003).
- Tkaczyk, C., Beaven, M. A., Brachman, S. M., Metcalfe, D. D. & Gilfillan, A. M. The phospholipase C γ 1-dependent pathway of Fc ϵ RI-mediated mast cell activation is regulated independently of phosphatidylinositol 3-kinase. *J. Biol. Chem.* **278**, 48474–48484 (2003).
- Fukao, T., Terauchi, Y., Kadowaki, T. & Koyasu, S. Role of phosphoinositide 3-kinase signaling in mast cells: new insights from knockout mouse studies. *J. Mol. Med.* **81**, 524–535 (2003).
- Campbell, E., Hogaboam, C., Lincoln, P. & Lukacs, N. W. Stem cell factor-induced airway hyperreactivity in allergic and normal mice. *Am. J. Pathol.* **154**, 1259–1265 (1999).
- Hundley, T. R. *et al.* Kit and Fc ϵ RI mediate unique and convergent signals for release of inflammatory mediators from human mast cells. *Blood* **104**, 2403–2417 (2004).
- Strait, R. T., Morris, S. C., Yang, M., Qu, X. W. & Finkelman, F. D. Pathways of anaphylaxis in the mouse. *J. Allergy Clin. Immunol.* **109**, 658–668 (2002).
- Fukao, T. *et al.* Selective loss of gastrointestinal mast cells and impaired immunity in PI3K-deficient mice. *Nature Immunol.* **3**, 295–304 (2002).
- Chen, D. *et al.* p500/p55 α phosphoinositide 3-kinase knockout mice exhibit enhanced insulin sensitivity. *Mol. Cell. Biol.* **24**, 320–329 (2004).
- Ueki, K. *et al.* Increased insulin sensitivity in mice lacking p85 β subunit of phosphoinositide 3-kinase. *Proc. Natl Acad. Sci. USA* **99**, 419–424 (2002).
- Terauchi, Y. *et al.* Increased insulin sensitivity and hypoglycaemia in mice lacking the p85 α subunit of phosphoinositide 3-kinase. *Nature Genet.* **21**, 230–235 (1999).
- Mauvais-Jarvis, F. *et al.* Reduced expression of the murine p85 α subunit of phosphoinositide 3-kinase improves insulin signaling and ameliorates diabetes. *J. Clin. Invest.* **109**, 141–149 (2002).
- Fruman, D. A. *et al.* Hypoglycaemia, liver necrosis and perinatal death in mice lacking all isoforms of phosphoinositide 3-kinase p85 α . *Nature Genet.* **26**, 379–382 (2000).
- Wymann, M. P., Zvelebil, M. & Laffargue, M. Phosphoinositide 3-kinase signalling—which way to target? *Trends Pharmacol. Sci.* **24**, 366–376 (2003).

- Gray, A., Olsson, H., Batty, I. H., Priganica, L. & Peter Downes, C. Nonradioactive methods for the assay of phosphoinositide 3-kinases and phosphoinositide phosphatases and selective detection of signaling lipids in cell and tissue extracts. *Anal. Biochem.* **313**, 234–245 (2003).
- Okayama, Y., Tkaczyk, C., Metcalfe, D. D. & Gilfillan, A. M. Comparison of Fc ϵ RI- and Fc γ RI-mediated degranulation and TNF- α synthesis in human mast cells: selective utilization of phosphatidylinositol-3-kinase for Fc γ RI-induced degranulation. *Eur. J. Immunol.* **33**, 1450–1459 (2003).
- Tilley, S. L., Wagoner, V. A., Salvatore, C. A., Jacobson, M. A. & Koller, B. H. Adenosine and inosine increase cutaneous vasopermeability by activating A(3) receptors on mast cells. *J. Clin. Invest.* **105**, 361–367 (2000).
- Laffargue, M. *et al.* Phosphoinositide 3-kinase gamma is an essential amplifier of mast cell function. *Immunity* **16**, 441–451 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Profit and G. Willams for bioanalytical support. K.A. and A.B. were supported in part by an MRC-CASE and Fondazione Italiana per la Ricerca sul Cancro fellowship, respectively. Research in the laboratory of B.V. is funded by the Ludwig Institute for Cancer Research, the Biotechnology and Biological Science Research Council and European Union FP5 and FP6 Programmes.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.V. (bartvanh@ludwig.ucl.ac.uk).

DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1

Grzegorz Ira^{1*}, Achille Pelliccioli^{2*}, Alitukiriza Balijja², Xuan Wang^{1†}, Simona Fiorani², Walter Carotenuto², Giordano Liberi², Debra Bressan¹, Lihong Wan³, Nancy M. Hollingsworth³, James E. Haber¹ & Marco Foiani²

¹Rosenstiel Center and Department of Biology, Brandeis University, Waltham, Massachusetts 02454-9110, USA

²F.I.R.C. Institute of Molecular Oncology Foundation, Via Adamello 16, 20139, Milano, Italy, and Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Italy

³Department of Biochemistry and Cell Biology, SUNY Stony Brook, Stony Brook, New York 11794-5215, USA

* These authors contributed equally to this work

† Present address: Rockefeller University, 1230 York Avenue, New York, New York 10021-6399, USA

A single double-strand break (DSB) induced by HO endonuclease triggers both repair by homologous recombination and activation of the Mec1-dependent DNA damage checkpoint in budding yeast^{1–6}. Here we report that DNA damage checkpoint activation by a DSB requires the cyclin-dependent kinase CDK1 (Cdc28) in budding yeast. CDK1 is also required for DSB-induced homologous recombination at any cell cycle stage. Inhibition of homologous recombination by using an analogue-sensitive CDK1 protein^{7,8} results in a compensatory increase in non-homologous end joining. CDK1 is required for efficient 5' to 3' resection of DSB ends and for the recruitment of both the single-stranded DNA-binding complex, RPA, and the Rad51 recombination protein. In contrast, Mre11 protein, part of the MRX complex, accumulates at unresected DSB ends. CDK1 is not required when the DNA damage checkpoint is initiated by lesions that are processed by nucleotide excision repair. Maintenance of the DSB-induced checkpoint requires continuing CDK1 activity that ensures continuing end resection. CDK1 is also important for a later step in homologous recombination, after strand invasion and before the initiation of new DNA synthesis.

In budding yeast, a chromosomal DSB created by HO endonuclease has been used both to study the kinetics and efficiency of DSB

repair and to analyse the induction of the DNA damage checkpoint dependent on Mec1 (an ATR homologue). In cells carrying *HML* or *HMR* mating-type switching donor sequences, a DSB at the *MAT* locus is efficiently repaired by gene conversion. In strains lacking donor sequences, induction of an unreparable DSB causes arrest of cell cycle progression before anaphase^{1,2}. In both instances, a key step is the 5' to 3' resection of DSB ends to produce single-stranded DNA (ssDNA), which is bound by the RPA complex. RPA binding is essential both for association of Mec1 checkpoint kinase⁹ and for loading of Rad51 recombination protein⁶.

Activation of the Mec1-dependent DNA damage checkpoint after a DSB is regulated by the cell cycle³, with no activation in G1-arrested cells. A DSB induced in cells that have been arrested in G1, and then released into S phase, results in hyperphosphorylation of the Mec1 target Rad53 after the completion of S phase, in G2 (Supplementary Fig. S1a). To test whether the checkpoint depends on the activity of cyclin-dependent kinases, we inactivated CDK1 in nocodazole-blocked G2 cells. We overexpressed the CDK1/Clb inhibitor, Sic1 (ref. 10), in G2 cells at the same time that an unreparable DSB was induced at *MAT*. CDK1 inactivation is measured by the progressive dephosphorylation of the B subunit of DNA polymerase- α , a marker for Cdc28/Clb activity¹¹ (Fig. 1). *SIC1* overexpression prevents the accumulation of phosphorylated Rad53 and Chk1 (Fig. 1) and impairs hyperphosphorylation of the upstream checkpoint factors Ddc2 and Rad9 as well as Mre11 (Fig. 1). Because the phosphorylation of Ddc2 and Rad9 is directly mediated by Mec1 kinase, we conclude that CDK1 inactivation affects Mec1.

To determine whether G1-arrested cells are able to perform homologous recombination (HR), we arrested *MATa* cells with α -factor and then induced a DSB with HO. Recombination between *HMLa* and *MATa*, monitored by Southern blots, was nearly absent in comparison with that in cells growing exponentially or in cells

arrested in G2 with nocodazole, where replacement of *MATa* by *MAT α* was efficient (Fig. 2a).

Inhibition of HR in G1-arrested cells was also seen in a diploid where a DSB at *MAT* could be repaired only by allelic recombination with an uncleavable *MATa-inc* locus on a homologous chromosome (Fig. 2e). We obtained similar inhibition of HO-induced recombination in G1 haploid cells when an HO-induced DSB was repaired by ectopic recombination between chromosomes III and V (ref. 5; data not shown).

To establish that the failure of HO-induced recombination in G1-arrested cells was attributable to a lack of Cdc28 activity, we examined recombination in cells expressing Cdc28-as1, a mutant sensitive to the ATP analogue inhibitor 1-NMPP1 (ref. 7). Galactose-induced HO cleavage was efficient, but cell-cycle-arrested cells failed almost completely to repair the DSB (Fig. 2b). 1-NMPP1 did not affect repair in wild-type cells (Fig. 2c).

CDK1 is also required for HR in the G2/M stage of cell cycle. We arrested *cdc28-as1* cells in G2 with nocodazole and then induced the expression of HO. Whereas recombination was normal in G2-arrested cells, *MAT* switching was nearly abolished in Cdc28-inhibited cells (Fig. 2b).

Failure of both checkpoint activation and HR in G1-arrested cells and in both Sic1-inhibited and Cdc28-as1-inhibited G2 cells correlates with an absence of 5' to 3' resection of DSB ends. The effect of overexpressing *GAL1::SIC1* in nocodazole-arrested G2 cells was shown by examining the rate of loss of the HO-cut *MATa EcoRI* restriction fragment in strains lacking *HML* or *HMR*, where the DSB is not repaired (Fig. 3a). Inhibiting CDK1 counteracted degradation of the HO-cut fragment as well as additional fragments a greater distance from the DSB. Resection was not significantly affected in G2-arrested cells deleted for Rad9, Rad17 or Mec1 checkpoint proteins¹² (data not shown).

Similar defects in resection were seen in α -factor-arrested G1 cells and in all stages of the cell cycle when Cdc28-as1 was inhibited (Fig. 3a). Without resection, HO-cleaved *MAT* sequences fail to bind either RPA or Rad51 (Fig. 3b and Supplementary Fig. S1) as determined by chromatin immunoprecipitation (ChIP). A similar failure of RPA loading was seen when CDK1 was inhibited by Sic1 overexpression (Fig. 3b). In contrast, both resection and RPA and Rad51 binding are seen in nocodazole-arrested G2 cells, in which CDK1 is active (Fig. 3b). Without RPA and Rad51 binding, HR should not occur. The absence of RPA recruitment to DSB ends in CDK1-inhibited cells also accounts for the failure to activate the Mec1-dependent DNA damage checkpoint, because Mec1-Ddc2 recruitment depends on prior binding of RPA⁹.

The 5' to 3' resection of HO-induced DSB ends is reduced, but not eliminated, in cells deleted for *MRE11*, *RAD50* or *XRS2* (ref. 1). However, there are cell cycle differences in the control of resection. In G2-arrested cells, 5' to 3' resection depends almost completely on the MRX complex¹³, but in G1-arrested cells there is still residual resection when *RAD50* is deleted (Fig. 3a). When Cdc28-as1 kinase was inhibited in wild-type G2-arrested cells, resection was as defective as in *rad50 Δ* G2-arrested cells (Fig. 3a). We conclude that Cdc28 controls resection activity associated with the MRX complex. We note that Cdc2 in *Schizosaccharomyces pombe* has been suggested to control the resection of DSBs, but independently of Rad50 (ref. 14). Without resection, *mre11 Δ* G2-arrested cells also fail to activate the DNA damage checkpoint¹⁵ (Fig. 3c).

Because Mre11 is required for DSB resection and checkpoint activation^{1,15,16}, we asked whether defective resection caused by CDK1 inactivation reflects the inability to load Mre11 on broken chromosomes. We found Mre11 at DSBs in G2 cells (Fig. 3b and Supplementary Fig. S1), even when CDK1 is inactive. However, whereas in wild-type cells the amount of Mre11 crosslinked near the DSB declines by 2 h, in Sic1-overexpressing cells Mre11 remains associated with unresected ends. Transient association of Mre11 with DSB ends in wild-type cells might suggest that MRX proteins

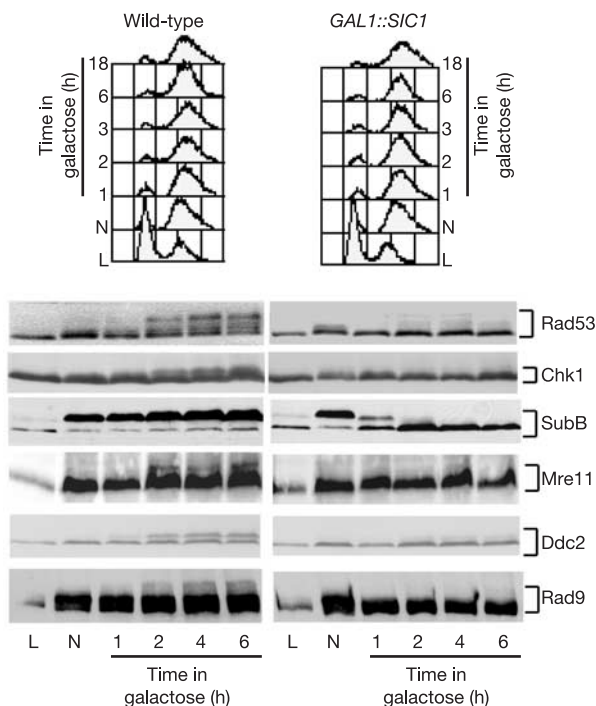


Figure 1 CDK1 activity is required for DSB-induced phosphorylation of checkpoint proteins in G2 cells. The phosphorylation of checkpoint proteins in the presence of an HO-induced unrepaired DSB in G2/M cells arrested with nocodazole (N) is shown, comparing cells with active CDK1 or *GAL1::SIC1*-inactivated CDK1.

migrate down DNA as resection proceeds; however, we could not detect significant association of Mre11 with DNA 5 kilobases (kb) from the HO cut even after several hours (when resection has reached and passed this region⁶) (data not shown). However, it is possible that Mre11 becomes spread out and undetectable. CDK1 inactivation therefore does not prevent Mre11 loading onto DSBs.

Although G1-arrested cells fail to perform efficient DSB-induced HR, they exhibit an increased efficiency of non-homologous end joining (NHEJ). In strains lacking *HML* or *HMR*, repair of the DSB

occurs only by Ku- and DNA ligase 4-dependent rejoining of the 4-base-pair (bp) 3'-overhanging complementary HO-cleaved ends¹⁷. When 1-NMPP1 was added 30 min before 1 h of HO induction in growing cells, NHEJ increased fivefold, from 12% to 65% (Fig. 2f). The same increase in NHEJ was observed when CDK1 inhibitor was added to nocodazole-arrested G2 cells (Fig. 2f). These results support and extend previous suggestions^{18,19} that the higher stability of DSB ends in G1-arrested cells stimulates NHEJ, which we propose might result from a lack of resection of these ends. In

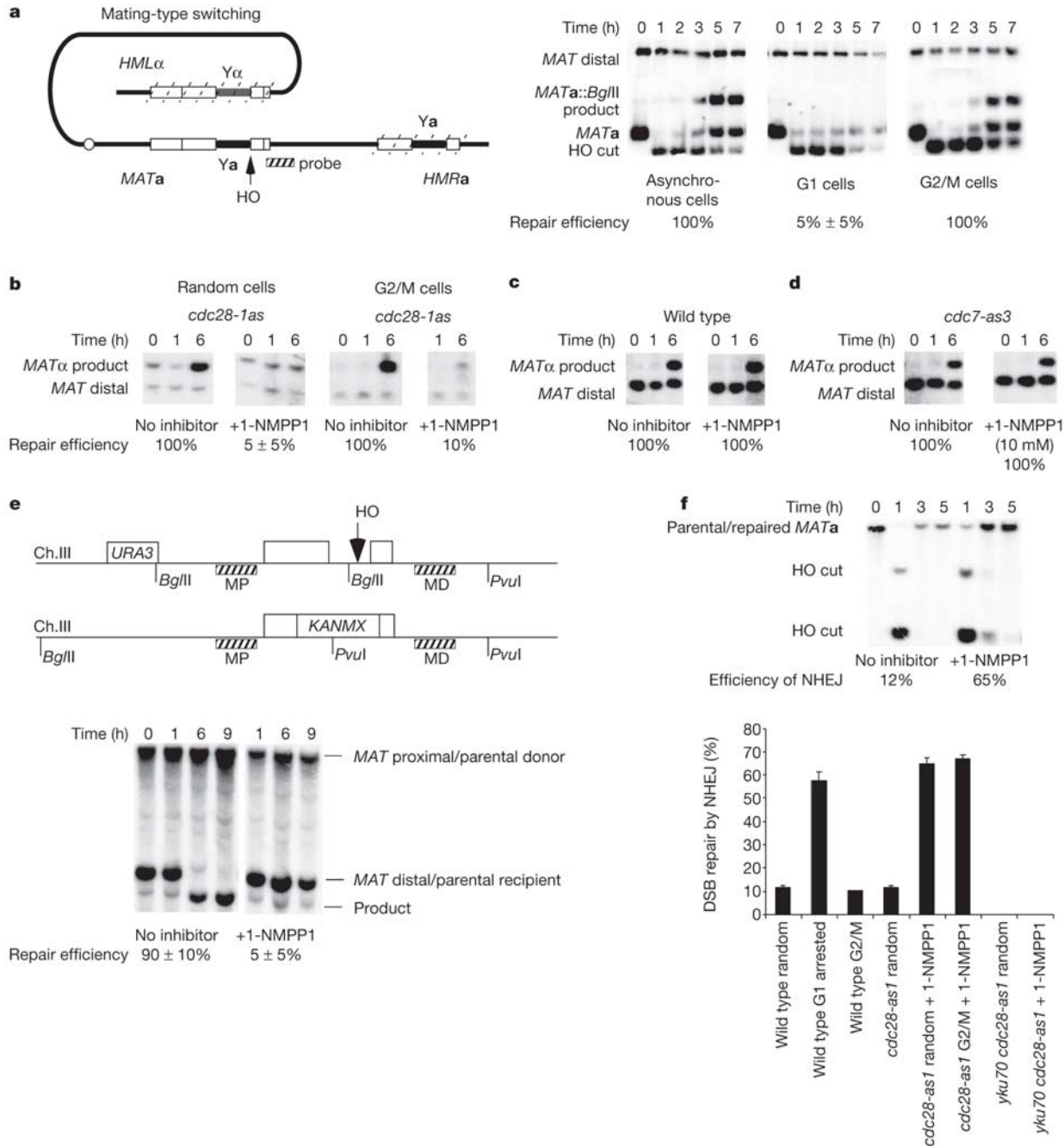


Figure 2 CDK1 is required for homologous recombination. **a**, *MAT* switching is initiated by creating an HO-induced DSB at the *MAT* locus that is repaired by gene conversion from *HML* or *HMR*. *MAT* switching is shown in asynchronous cells or cells arrested in G1 (α -factor arrest) or G2/M (nocodazole arrest). **b–d**, *MAT* switching in *cdc28-as1* (**b**), wild-type (**c**) and *cdc7-as3* (**d**) strains with and without 1-NMPP1 inhibitor. **e**, Allelic

recombination in a *cdc28-as1* homozygous diploid strain. The position of two DNA probes is shown: *MAT*-distal (MD, verifies DSB induction) and *MAT*-proximal (MP, detects product). **f**, Efficiency of NHEJ in *cdc28-as1* cells with either active or inactive CDK1 at different stages of the cell cycle. Error bars indicate s.d.

addition, NHEJ is probably aided by the retention of the end-joining protein Mre11 at DSB ends (Fig. 3b).

Lack of checkpoint activation in CDK1-inhibited cells proves to be specific for DSB damage and not for other conditions that induce Mec1-dependent checkpoint activation. For example, G1 cells exposed to the ultraviolet-mimetic agent 4-nitroquinoline 1-oxide (4NQO) exhibit Rad53 phosphorylation¹¹ (Fig. 4a). CDK1 activity is therefore not required when the checkpoint is activated by presumably ssDNA intermediates of nucleotide excision repair. Furthermore, CDK1 inhibition does not affect checkpoint activation in response to S-phase arrest caused by hydroxyurea²⁰, during which ssDNA intermediates arise at stalled replication forks²¹. We conclude that CDK1 influences Mec1-dependent checkpoint activation only when the DNA damage involves DSBs that must be resected to generate ssDNA.

We tested whether CDK1 activity was required for the maintenance of the active checkpoint state and continuing DSB resection.

DSB formation was induced in wild-type and *cdc28-as1* G2 cells. Both strains promote robust checkpoint activation, but when Cdc28-as1 was inactivated by adding 1-NMPP1, both the checkpoint and DSB resection were turned off (Fig. 4b, c). When inhibitor is added, a region 20 kb from the DSB fails to be degraded (Fig. 4c). Hence, CDK1 is required for ongoing resection and for maintaining the checkpoint. Indeed we believe that the signal for checkpoint maintenance is not simply the presence of ssDNA but rather some event that is linked to the continuing resection process.

Beyond affecting the formation of ssDNA, CDK1 seems to have a second function during HR. First, although in G2-arrested cells both the *rad50Δ* mutation and inhibition of Cdc28-as1 have the same effect on resection (Fig. 3a), recombination still occurs at about 20% of wild type in *rad50Δ* cells but is completely abolished when Cdc28-as1 is blocked (data not shown). This suggests that Cdc28 must affect other steps in DSB repair. To examine CDK1's role more directly, we added inhibitor to *cdc28-as1* cells at intervals

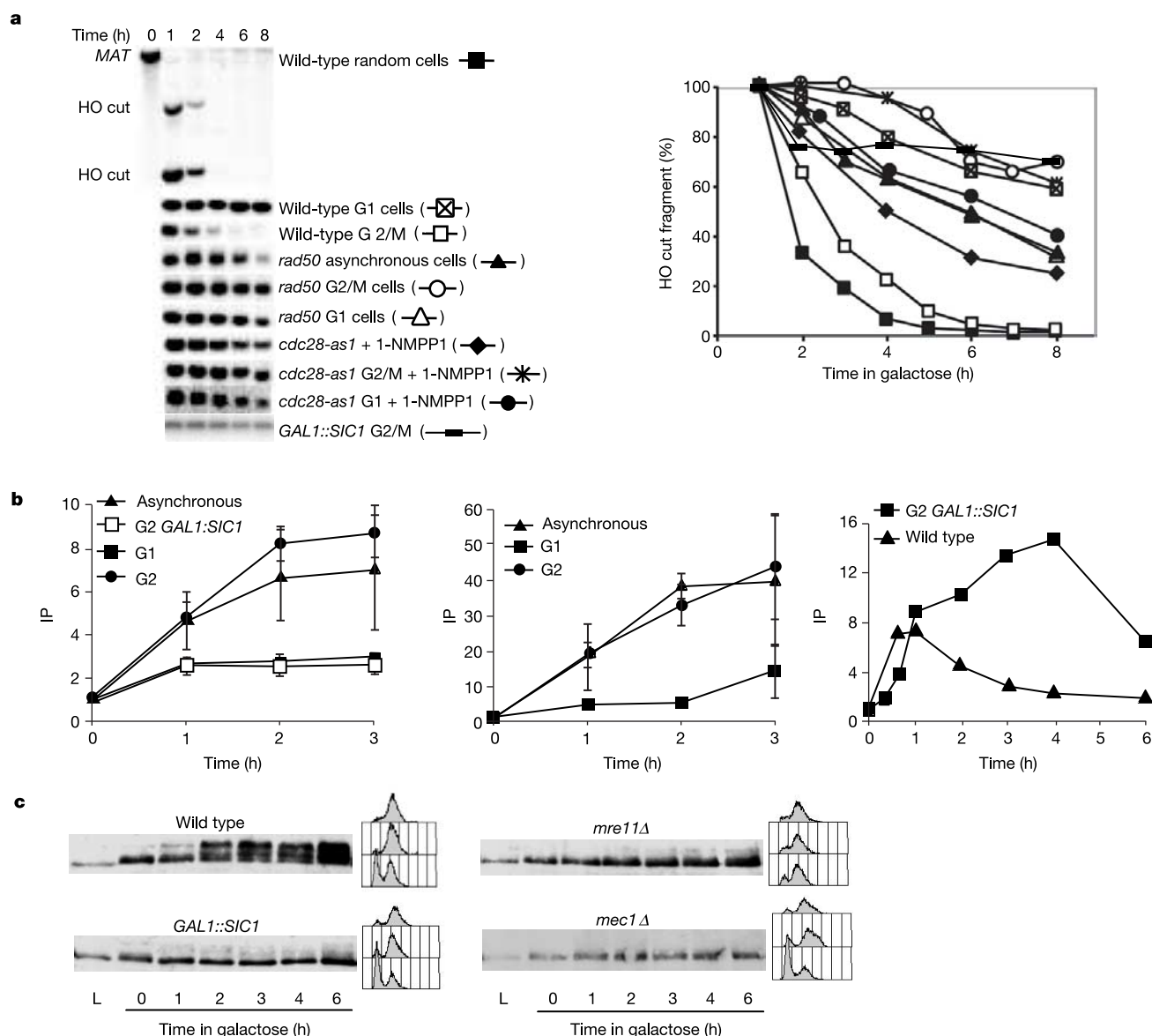


Figure 3 CDK1 is needed for DSB resection. **a**, The 5' to 3' resection of DSB ends was analysed in the indicated mutants and at different points in the cell cycle by loss of the HO-cut *EcoRI* restriction fragment in a strain lacking *HML* or *HMR*. **b**, ChIP analysis of binding of RPA (left), Rad51 (centre) and Mre11 (right) to DSB ends in a donorless strain^{4,6}.

in asynchronous or G1 cells and in G2/M-arrested cells with or without Sic1 inhibition of CDK1. Where error bars are shown, results are means $\pm$ s.d. **c**, Lack of Rad53 phosphorylation in *mre11Δ* G2/M-arrested cells, compared with other conditions.

after initiation of HO cleavage. When inhibitor was added 4–6 h after break induction, the extent of repair was unaffected, but at earlier times the addition of inhibitor reduced the completion of DSB repair (Fig. 5a). The most striking results were found when inhibitor was added 2 h after DSB induction. By this time, most DSB ends have been resected sufficiently to bind both Rad51 and RPA^{4,6}. This is evident from ChIP analyses showing that, by 90 min, RPA and Rad51 had bound near the HO cut as efficiently as in wild-type cells. Moreover, Rad51-ssDNA filaments succeeded in strand invasion and synapsis with the donor, so that the *HML* locus (200 kb from *MAT*) could be immunoprecipitated by anti-Rad51 antibody⁴. As shown in Fig. 5b and Supplementary Fig. S1, Rad51-mediated synapsis between *MAT* and *HML* was comparable to that seen without inhibitor at each time up to 4 h, when *MAT* switching is complete without inhibitor. However, the next step in repair, the initiation of new DNA synthesis (primer extension) from the 3' end of the invading strand, was severely impaired in cells to which inhibitor was added (Fig. 5c and Supplementary Fig. S1). We conclude that CDK1 is required after synapsis and before the initiation of new DNA synthesis, a process that requires the loading of the proliferating cell nuclear antigen (PCNA) clamp and the recruitment of DNA polymerase.

The Cdc7 protein kinase is activated by CDK1 and is required for the initiation of DNA replication²². Using an analogue-sensitive allele of *CDC7*, *cdc7-as3*, we showed that HO-induced recombination is efficient in *Cdc7-as3*-inhibited cells (Fig. 2d). Thus, cells arrested in late G1, before the initiation of S phase but beyond the START point where CDK1 is activated, are capable of resecting DSB ends and completing HR.

Results presented here reveal a central role for budding yeast's CDK1, Cdc28, in the regulation of DSB repair. First, when CDK1 is inactive, in G1 cells before passing START, HR is markedly reduced, as it is in G1-stage mammalian cells. This observation is supported by the finding that G1-arrested yeast cells fail to form DNA-damage-induced foci of Rad52 protein²³. Moreover, the DNA damage checkpoint, which requires extensive resection of DSB ends, is not activated in G1-arrested cells³. We note that HR is not completely abolished in G1-arrested or *Cdc28*-inhibited cells; there is still

residual 5' to 3' resection of DSB ends (apparently by a Cdc28-independent exonuclease) and about 5–10% of product is still formed. The CDK1-dependent regulation of HR and NHEJ might also help to explain the cell-cycle specificity of DSB repair in mammalian cells. Mammalian G1 cells predominantly repair ionizing radiation damage by NHEJ²⁴. Similarly, V(D)J recombination by NHEJ during lymphoid development is restricted to G0/G1 phases of the cell cycle by the tight regulation of *RAG2* gene expression and protein degradation. When *RAG2* is expressed throughout the cell cycle it causes aberrant recombination products involving HR²⁵. Once cells have passed START, HR is activated, allowing DSBs arising during replication to be repaired efficiently by HR.

We show that CDK1 is required for the establishment and maintenance of the Mec1/Rad53-mediated checkpoint response after DSB formation and for initiating and sustaining Mre11-dependent DSB resection. CDK1-mediated DSB resection influences checkpoint activation by allowing the formation of RPA-ssDNA filaments. Mre11 loading at DSB ends is not in itself sufficient to promote checkpoint activation, suggesting that Mec1-dependent checkpoint activation is instead influenced by the Mre11-mediated formation of RPA filaments. A tantalizing possibility is that in G2 CDK1 regulates the phosphorylation state of the MRX complex and/or other as yet unidentified factors implicated in DSB processing. Indeed several proteins implicated in DNA recombination are targeted by CDKs, including Xrs2, RPA and Srs2 (refs 8, 20). However, we note that neither *mre11-T294V/S379A/S413A/T529V/S560A/T677V* nor *xrs2-T108V/S542A* (in which the putative Cdc28-dependent phosphorylation sites have been mutagenized) exhibit hypersensitivity to DNA-damaging agents or appreciable defects in checkpoint activation after a DSB in G2 (data not shown). CDKs have been also implicated in the phosphorylation of checkpoint proteins (namely Rad9, Drc1/Sld2, Ddc1, Dun1 and Ptc3)^{8,26,27}; however, it is not known whether these phosphorylation events have any role in checkpoint activation or rather in recovery from damage.

A hypothesis suggested by our data is that CDK1 has a key function in regulating alternative DSB repair pathways (NHEJ or HR) in G1 and G2. Hence, CDK1 inactivation in G2 seems to

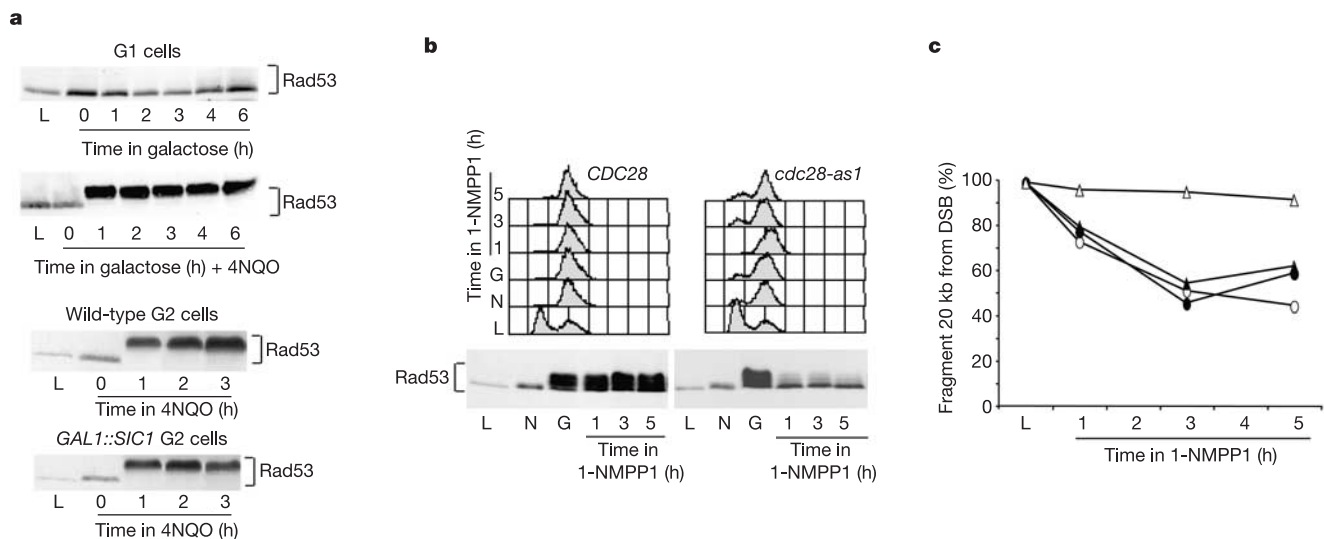


Figure 4 CDK1 is required to maintain Rad53 activation in response to a DSB but not after damage by 4NQO. **a**, In G1-arrested cells, Rad53 is not phosphorylated when an HO-mediated DSB is induced by galactose, but is triggered when 4NQO provokes nucleotide excision repair. In nocodazole-arrested G2/M cells, treatment with 4NQO causes Rad53 phosphorylation even when CDK1 is inhibited by Sic1. **b**, At 6 h after DSB induction, when

the checkpoint is established and Rad53 is phosphorylated, 1-NMPP1 was added to either *CDC28* or *cdc28-as1* strains. **c**, Rad53 phosphorylation is lost when CDK1 is inactivated, and resection of a *StyI* fragment, 20 kb from the DSB, is inhibited. Open triangles, *cdc28-as1* G2 + 1-NMPP1, filled triangles, *cdc28-as1* G2, open circles, wild type G2 + 1-NMPP1.

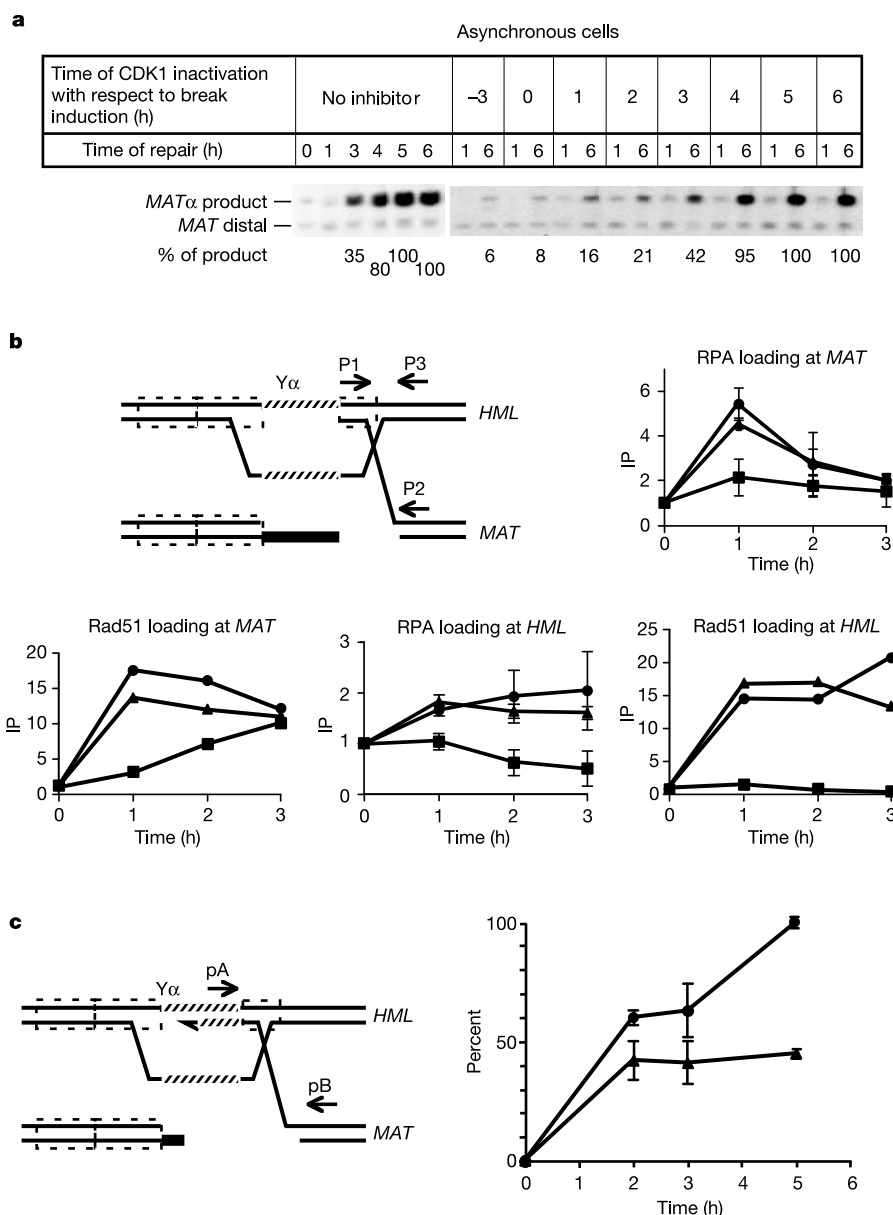


Figure 5 Role of *CDK1* in a late stage of *MAT* switching. **a**, *Cdc28-as1* was inhibited in cells at different times before and after *HO* induction, and the amount of product formed at 6 h was determined. **b**, Strand invasion was tested by ChIP analysis for RPA and Rad51 in a *cdc28-as1* strain. The diagram shows homologous sequences (dashed outlined boxes) and primers used to detect immunoprecipitation (IP) of *MAT* (p1, p2) or *HML* (p1, p3) with

anti-Rad51 or Rfa1 antibodies. Symbols indicate the time at which inhibitor was added with respect to break induction: triangles, no inhibitor; squares, 0.5 h; circles, 1.5 h. **c**, Polymerase chain reaction detects the first new DNA synthesis primed by the 3' invading end after strand invasion. Circles, no inhibitor; triangles, 1-NMPP1 added 1.5 h after *HO* induction. Where error bars are shown, results are means $\pm$ s.d.

promote a G1-like repair response. Since Mre11 is required both for NHEJ and HR²⁸, perhaps in CDK1-limiting G2 cells, Mre11 is 'frozen' in a NHEJ-proficient status, unable to resect DSB ends and therefore to generate checkpoint signals. This is also in accordance with observations *in vitro* showing that MRX complexes protect DNA ends from other exonucleases²⁹. □

Methods

Strains

Strains used to study DSB ends resection, DNA damage checkpoint activation and NHEJ were derivatives of JKM139 (*hoΔ MATα hmlΔ::ADE1 hmrΔ::ADE1 ade1-100 leu2-3,112 lys5 trp1Δ::hisG ura3-52 ade3::GAL::HO*) or JKM179 (same genotype except *MATα*), carrying *GAL1::SIC1* (integrating pMHT plasmid with a stable version of the *SIC1* gene under the *GAL1* promoter¹¹, kindly provided by J. Diffley); *CHK1-13MYC*; *DDC-4HA*; *RAD9-13MYC*; *mre11Δ*; *mec1Δsml1Δ*; *rad50Δ*; *bar1Δ*; *cdc28-as1*; *cdc5-ad* or the combination of these mutations. In *MATα-inc* strains the *HO* recognition site has a 1-bp substitution that prevents *HO* endonuclease cleavage. The diploid strain used to study

DSB-induced checkpoint activation and resection was *MATα/MATα*-like, where part of the *Yα* sequence was replaced with *KANMX*. *MAT* switching was tested in *MATα HMLα HMRα ade3::GAL::HO ade1-100 leu2-3,112 lys5 ura3-52 bar1Δ* strains, some carrying *cdc28-as1* or *cdc7-as3*. In the experiment shown in Fig. 2b, gene conversion from both *HML* and *HMR* yields cells that remain α -factor sensitive because both *HMR* and *HML* contain *Yα* information. In the strain used to study allelic recombination, the recipient chromosome has an insertion of *URA3* at *PHO87* located 5 kb to the left of *MAT*, whereas the donor chromosome has the *Yα* and *Z1* sequences of *MATα* replaced by *KANMX* and is not cut by *HO* endonuclease.

Fluorescence-activated cell sorting analysis

Fluorescence-activated cell sorting was performed as described³.

CDK1 inactivation, α -factor arrest and nocodazole arrest

CDK1 and Cdc7 were inactivated in *cdc28-as1* and *cdc7-as3* strains by adding the ATP analogue inhibitor 1-NMPP1 at 5 μ M unless otherwise indicated. In *cdc28-as1*, *cdc7-as3* and wild-type cells, inhibitor was added 1.25 doubling times before inducing *HO* unless otherwise indicated. Cells were arrested in G1 with 10 μ g ml⁻¹ α -factor (or 0.5 μ g ml⁻¹ in *bar1Δ* strains). Cells were arrested in G2/M with 20 μ g ml⁻¹ nocodazole.

HO induction

Strains were grown in YP-lactate (1% Bacto yeast extract, 2% Bacto peptone, 3% lactic acid, pH 5.5) and HO induction was done as described previously⁴.

5'-3' strand resection at a DSB

Resection was measured as a rate of HO-cut band disappearance. In some cases the resection of sequences distant from the break was also measured. Purified genomic DNA, digested with *EcoRI* (Fig. 3) or *StyI* (Fig. 4), was separated on a 1% agarose gel, transferred to Hybond N⁺ and probed with ³²P-labelled DNA from *MATa* and from sequences 20 kb proximal to the DSB to establish the rate of resection. Hybridization to *HIS4* or *LEU2* (both more than 100 kb away) was used to normalize the amount of DNA at each time point. The rate of resection at each time point was plotted as the percentage of the density of the initial cut band. The density of the HO-cut band at *t* = 1 h was set to 100%.

NHEJ efficiency

NHEJ was examined in a donorless *cdc28-as1* strain. 1-NMPP1 inhibitor was added 30 min before HO induction. Re-cutting of *MATa* by HO was prevented by filtering cells out of galactose-containing medium 30 min after DSB induction and diluting cells into YP-dextrose. (1% Bacto yeast extract, 2% Bacto peptone, 2% dextrose, pH 5.5). The efficiency of NHEJ was determined as the intensity of the *MATa*-containing restriction fragment 3 h after HO induction, normalized to the amount of DNA.

Analysis of homologous recombination

MAT switching and other homologous recombination events were analysed on Southern blots⁴. For *MAT* switching, genomic DNA was digested with *StyI* and *BglII* (Fig. 2a) or with *StyI* (Fig. 2b–d) and probed with a *MAT*-distal probe. For allelic recombination, DNA was digested with *BglII* and *PvuI* and probed with the *MAT*-distal probe to check the efficiency of HO induction and with a *MAT*-proximal probe to check the appearance of the product. Initial new DNA synthesis after strand invasion was determined by polymerase chain reaction as described⁴, normalized to the amount of ARG5,6 DNA.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed as described previously^{4,6}. Antibodies against Mre11, Rad51 and RPA were provided by J. H. J. Petrini, A. Shinohara and S. Brill, respectively.

Western blots

Protein extracts were performed as described³. Antibodies used for western blots were Rad53 polyclonal antibody JD47 (a gift from J. Diffley), Mre11 polyclonal antibody (produced by the IFOM antibody facility) and monoclonal antibodies 9E10 (anti-Myc epitope), 12CA5 (anti-HA (haemagglutinin) epitope) and 6D2 (ref. 11) (anti-B subunit).

Received 1 June; accepted 19 August 2004; doi:10.1038/nature02964.

- Lee, S. E. *et al.* *Saccharomyces* Ku70, Mre11/Rad50 and RPA proteins regulate adaptation to G2/M arrest DNA damage. *Cell* **94**, 399–409 (1998).
- Toczyski, D. P., Galgoczy, D. J. & Hartwell, L. H. CDC5 and CKII control adaptation to the yeast DNA damage checkpoint. *Cell* **90**, 1097–1106 (1997).
- Pelliccioli, A., Lee, S. E., Lucca, C., Foiani, M. & Haber, J. E. Regulation of *Saccharomyces* Rad53 checkpoint kinase during adaptation from G2/M arrest. *Mol. Cell* **7**, 293–300 (2001).
- Sugawara, N., Wang, X. & Haber, J. E. *In vivo* roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol. Cell* **12**, 209–219 (2003).
- Vaze, M. B. *et al.* Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. *Mol. Cell* **10**, 373–385 (2002).
- Wang, X. & Haber, J. E. Role of *Saccharomyces* single-stranded DNA-binding protein RPA in the strand invasion step of double-strand break repair. *PLoS Biol.* **2**, 104–111 (2004).
- Bishop, A. C. *et al.* A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407**, 395–401 (2000).
- Ubersax, J. A. *et al.* Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859–864 (2003).
- Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
- Nasmyth, K. At the heart of the budding yeast cell cycle. *Trends Genet.* **12**, 405–412 (1996).
- Pelliccioli, A. *et al.* Activation of Rad53 kinase in response to DNA damage and its effect in modulating phosphorylation of the lagging strand DNA polymerase. *EMBO J.* **18**, 6561–6572 (1999).
- Naiki, T., Wakayama, T., Nakada, D., Matsumoto, K. & Sugimoto, K. Association of Rad9 with double-strand breaks through a Mec1-dependent mechanism. *Mol. Cell. Biol.* **24**, 3277–3285 (2004).
- Diede, S. J. & Gottschling, D. E. Exonuclease activity is required for sequence addition and Cdc13p loading at a *de novo* telomere. *Curr. Biol.* **11**, 1336–1340 (2001).
- Caspari, T., Murray, J. M. & Carr, A. M. Cdc2-cyclin B kinase activity links Crb2 and Rqh1-topoisomerase III. *Genes Dev.* **16**, 1195–1208 (2002).
- Grenon, M., Gilbert, C. & Lowndes, N. F. Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. *Nature Cell Biol.* **3**, 844–847 (2001).
- D'Amours, D. & Jackson, S. P. The yeast Xrs2 complex functions in S phase checkpoint regulation. *Genes Dev.* **15**, 2238–2249 (2001).
- Moore, J. K. & Haber, J. E. Cell cycle and genetic requirements of two pathways of nonhomologous

end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 2164–2173 (1996).

- Frank-Vaillant, M. & Marcand, S. Transient stability of DNA ends allows nonhomologous end joining to precede homologous recombination. *Mol. Cell* **10**, 1189–1199 (2002).
- Karathanasis, E. & Wilson, T. E. Enhancement of *Saccharomyces cerevisiae* end-joining efficiency by cell growth stage but not by impairment of recombination. *Genetics* **161**, 1015–1027 (2002).
- Liberi, G. *et al.* Srs2 DNA helicase is involved in checkpoint response and its regulation requires a functional Mec1-dependent pathway and Cdk1 activity. *EMBO J.* **19**, 5027–5038 (2000).
- Sogo, J. M., Lopes, M. & Foiani, M. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* **297**, 599–602 (2002).
- Foiani, M., Liberi, G., Lucchini, G. & Plevani, P. Cell cycle-dependent phosphorylation and dephosphorylation of the yeast DNA polymerase α -primase B subunit. *Mol. Cell. Biol.* **15**, 883–891 (1995).
- Lisby, M., Mortensen, U. H. & Rothstein, R. Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. *Nature Cell Biol.* **5**, 572–577 (2003).
- Rothkamm, K., Kruger, I., Thompson, L. H. & Lobrich, M. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol. Cell. Biol.* **23**, 5706–5715 (2003).
- Lee, J. & Desiderio, S. Cyclin A/CDK2 regulates V(D)J recombination by coordinating RAG-2 accumulation and DNA repair. *Immunity* **11**, 771–781 (1999).
- Toh, G. W. & Lowndes, N. F. Role of the *Saccharomyces cerevisiae* Rad9 protein in sensing and responding to DNA damage. *Biochem. Soc. Trans.* **31**, 242–246 (2003).
- Masumoto, H., Muramatsu, S., Kamimura, Y. & Araki, H. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature* **415**, 651–655 (2002).
- D'Amours, D. & Jackson, S. P. The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Rev. Mol. Cell Biol.* **3**, 317–327 (2002).
- Chen, L., Trujillo, K., Ramos, W., Sung, P. & Tomkinson, A. E. Promotion of Dn14-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. *Mol. Cell* **8**, 1105–1115 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The order of the co-first authors (G.I. and A.P.) and the order of the second two authors (A.B. and X.W.) were decided alphabetically. A.B. and X.W. made substantial and equal contributions to this work. We thank C. Zhang and K. Shokat (supported by NIH grant AI44009) for the *cdc28-as1* allele and the 1-NMPP1 inhibitor. We note the ownership of the ASKA technology by Cellular Genomics Inc. and thank CGI for granting us the rights to perform the experiments discussed in the publication. We thank J. Petrini, J. Diffley, K. Labib, M. P. Longhese, S. Brill and the IFOM antibody facility for reagents. We are grateful to C. Lucca, E. Pandiani, M. B. Vaze, S. Francia, S. Lovett, M. Lichten and the members of our laboratories for discussions and technical support. N.M.H. was supported by the NIH. M.F. and A.P. were supported by grants from Associazione Italiana per la Ricerca sul Cancro and M.F. from Telethon-Italy and the European Community. G.I. was supported by Charles A. King Trust Postdoctoral Fellowship. G.I. is on leave from N. Copernicus University, Torun, Poland. J.E.H. was supported by the NIH and the DOE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F. (marco.foiani@ifom-ieo-campus.it) or J.E.H. (haber@brandeis.edu).

erratum

High-resolution structure of a retroviral capsid hexameric amino-terminal domain

Guinahar B. Mortuza, Lesley F. Haire, Anthony Stevens, Stephen J. Smerdon, Jonathan P. Stoye & Ian A. Taylor

Nature **431**, 481–485 (2004).

In this Letter, the following sentence should have appeared after the author's email address: 'Coordinates and structure factors have been deposited in the Protein Data Bank under accession number 1U7K.' □

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

**Pennsylvania, New York,
Maryland/ Latin America/ NIH:**

Michael McGillion

Midwest USA:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

e-mail: p.phelan@nature.com

o.wulffen@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: r.asami@naturejpn.com

naturejobs

Rewarding season

The fall brings a flurry of awards for scientific research. The high-profile honours recognize well-established researchers, and many of the recipients of one eventually go on to win another — a sizeable subset of Lasker Award winners have subsequently received Nobel Prizes, for example.

One could say, tongue in cheek, that those people don't really need recognition at this point in their career. After all, they tend to have good funding and facilities already. Sure, the award money is nice. And statues on one's desk convey a certain gravitas.

But it is younger researchers who really need the recognition that prizes bring. They tend to be doing much of the bench work themselves and often have to scramble for support to keep their lab running, even with a good publication record. Fortunately, such awards do exist, although they are less well known. A new prize, the European Young Investigator Awards (EURYI), gave out its first crop of €1.25-million (US\$1.55-million), five-year grants this summer.

Another, the EMBO Gold Medal, was awarded last week to María Blasco, director of the molecular oncology programme at the National Cancer Centre in Madrid. She is the first Spanish citizen and only the third woman to receive this award for young European scientists since its launch in 1986. Individual countries are also following the trend, with Science Foundation Ireland's President of Ireland Young Researcher Award (PIYRA) being awarded last week to four young researchers, each of whom will receive up to €1.2 million over a five-year period.

Hopefully, more awards programmes like EURYI, PIYRA and the 'Gold' will follow. And they will aid as well as predict the future winners of those high-profile prizes.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

Finland on the move p1020

CAREER VIEW

Recruiters & Academia

Life changes

Postdoc Journal

Bridging the gap

Movers

Silke Bühler-Paschen p1022

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS



Harbouring potential:
Helsinki is trying to turn its
academic excellence into
commercial success.

Boosting biotech in Finland

Finland's science-funding situation seems enviable. This Nordic country has the second-highest per capita commitment to research of the 30 countries in the Organisation for Economic Co-operation and Development. It is second only to Sweden, with 3.5% of gross domestic product (GDP) going to research and development. Finland has a national venture-capital fund, to support start-up companies. And it has a public contract-research organization to provide expertise to both public and private institutions.

But this small country, with its big commitment to research and development, is vulnerable in some respects — as its scientific leaders will cheerfully admit. Most of its research and development funding comes from the private sector, with 70% coming from the cellphone manufacturer Nokia. Public-sector funding is tipped almost two to one in favour of applied research. And career obstacles — such as the lack of a well-developed postdoc programme, a reliance on short-term funding for graduate students and the absence of a tenure track in universities — may limit the attractiveness of the country to both nationals and foreigners.

But Finnish science leaders are rallying to address these challenges. In the public sector, the country is

equal to the United States and ahead of many western European countries in its funding of large multidisciplinary institutes, such as those around centres of academic excellence in Helsinki, Oulu and Turku. And for the private sector, the emphasis on applied research and government-supported venture capital has already created, if not another Nokia, a few mid-sized life-sciences companies.

PUBLIC CHALLENGES

Raimo Väyrynen, president of the Academy of Finland, which provides basic research funding, says that the present problems pale in comparison with those of the early 1990s, when Finland lost a greater percentage of its GDP than the United States did during the 1930s depression.

The government realized that it needed to lessen its reliance on agriculture, forestry and fishing, and invest in high-tech enterprises. The budget for research and development investment began the climb from 1.5% of GDP to its present position, and helped to turn Nokia into one of the country's biggest employers and a bastion of research and development spending.

Väyrynen suspects that the way to encourage another Nokia to emerge is to shift the country's funding from applied towards basic. Väyrynen says he will push the academy's €250-million (\$310-million) budget up by 7% each year over the next several years. The national technology agency, Tekes, which gives out the country's applied research funding, currently has a budget of about €400 million.

But "money alone is not the answer", says Väyrynen. The lack of a tenure-track system for senior scientists and the emphasis on short-term appointments for younger ones need to be addressed. He would also like to see Finland's research community become more international. Only 4.5% of

the country's scholars are from abroad, even though English is the country's working scientific language.

There are signs that infrastructure can turn that around, says Olli Jänne, director of Biocentrum at the University of Helsinki. About 15% of the building's 1,200 or so occupants speak neither Finnish nor Swedish and Jänne thinks he can increase that. His goal is to boost the research staff by 20% over the next few years, ideally with half of all new scientists coming from outside the country. "I would like to make this place more attractive to postdoctoral fellows from western Europe and the United States," says Jänne.

The building itself could be a big draw. The 24,000-square-metre facility is designed to maximize serendipitous interaction. Biocentrum's labs are clearly visible from an interior courtyard, and spiral staircases lead to a cafeteria on the ground floor. It also has the typically Finnish faculty lounge with sauna on the top floor, available for special events.

The building has had some success in encouraging interactions. There are already two successful spin-off companies: systems-biology firm MediCel and diagnostics company MoBiDiag. Two wings for biotech incubators are being added and additional lab space will be available in the complex, which is linked by tunnels to nearby hospitals, in 2006.

North of Helsinki, about 200 kilometres south of the Arctic Circle, an equally stunning building, opened at the end of 2003, may draw international researchers. Biocenter Oulu, an umbrella organization within the University of Oulu, supports 15 groups with a total research staff of 270. Taina Pihlajaniemi, scientific director of Biocenter Oulu, wants to add one or two groups a year over the next few years, to capitalize on the centre's main research themes — systems biology and intercellular interactions. These areas are coming to the fore in the global biology community.

And west of Helsinki, the University of Turku is adding a new positron emission tomography (PET) centre devoted entirely to research. This is unusual, as most medical research centres with PET facilities have to share them with clinicians. The Turku centre, in collaboration with UK-based Amersham Biosciences (now part of GE Healthcare), will help the university to study how drugs travel through the body.

This focus on pharmacodynamics is typical of the University of Turku, which has a history of supporting the life sciences. It fits well with the research themes of drug development and diagnostics, which are emphasized by BioTurku, an association of life-sciences companies and institutions in the city. Turku has the country's oldest science park, with about 600 researchers located in 11 buildings downtown, one for each discipline. There are about 40 bioscience companies in the Turku region, employing 3,000 people. Drug development and diagnostics, as well as the foodstuffs industry, form the core knowledge of the region, and serve as the basis for the new bioindustry.

Hormos Medical Corporation in Turku, a drug discovery and development company, is a textbook example of how the Finnish government nurtures biotech spin-offs. The company, founded in 1997 by University of Turku faculty and employees from Orion, Finland's sole homegrown big pharmaceutical company, has received €15 million over the years in loans and grants from Tekes, as well as start-up money from the Finnish National Fund for Research and Development (Sitra), the country's venture-capital arm. It also relied on VTT, a government-run contract-research organization, for outside research help. VTT recently opened an office in Turku and is filling it with 150 people.

Risto Lammintausta is Hormos's chief executive and is president of PharmaCluster Finland, an organization that aims to help biotech companies share resources. He says that Hormos probably could not have survived if it had had to rely on raising capital in the open market; it even had to shed some 25 employees in 2003. But now Lammintausta is optimistic that the company will grow again, as

its lead product, an oestrogen-replacement therapy, is entering phase III clinical trials.

Finnzymes, a Helsinki-based company specializing in restriction enzymes, illustrates how the government programmes that have helped Hormos to survive can also help a company achieve independence. Tekes had a "major" impact on the development of Finnzymes from a university spin-off into a fully fledged company, says founder and chief executive officer Pekka Mattila.

Tekes funded his work on collecting enzyme samples from geothermal hot springs, helping him to turn the enzymes into laboratory tools. An early marketing arrangement with New England Biolabs has helped the company stay solvent since its beginning, and VTT has aided Finnzymes by shouldering some of its production requirements.

Finnzymes is a major tenant in a new life-sciences park on the outskirts of Helsinki. Two buildings

are up, one more is under construction and two more are planned, as demand is high.

Companies such as Finnzymes and Hormos show that Finland holds promise in the life-sciences industry. But Timo Veromaa, of drug-development company BioTie Therapies in Turku, points out that growth of such companies is limited by Finland's lack of experienced researchers. The country has plenty of good scientists, but few who have seen a drug all the way through clinical trials. "Finland needs to attract people who have business experience, people who have pharma experience," says Veromaa. "That pool in Finland is too small."

But if Finland can marry its telecoms and information-technology successes with its biomedical promise — perhaps in diagnostics, biomedical engineering and medical imaging — then the country will have new ways of generating jobs, and new reasons to attract admiration.

Paul Smaglik is *Naturejobs* editor.



Oulu, famed for its policeman statue, has opened a centre for biological research.

POSTDOC JOURNAL

Bridging the gap

It's good to talk. While at Fermilab in Batavia, Illinois, I could pop along the corridor at any time and chat to a world expert about the topic of my choice — a real luxury. The hardest times of my PhD have been those when I felt most isolated and didn't talk to others. I guess I've really learned the benefits of bringing scientists together.

I don't just mean researchers from different fields, but also theorists and experimentalists from the same area. Different skills and a new perspective can be just what the doctor ordered when it comes to tackling the latest problems.

The most engaging debates I've had have been with scientists from the opposite end of the spectrum. My brother, a geneticist, asks me how to write programs to analyse his data. In return he shows me new statistical techniques that might be useful in my research. In my own field, the 'need to talk' has spurred some of us to organize a conference where young experimentalists and theorists will do exactly that — talk.

Multidisciplinary research centres are popping up all over the place. Bridging the gap between one field and another is risky for a scientist, but the rewards can be high. And as people recognize this untapped potential, interdisciplinary science can only go from strength to strength.

Amber Jenkins is a graduate student in particle physics at Imperial College London.

RECRUITERS & ACADEMIA

My job as a senior grants administrator — heading the pre-award unit within the research support office — is a secure one. But it is also a dead-end position.

In order to increase my skill set and plan for the future, I enrolled in the doctorate programme at my university's department of information science. I knew that the skills I picked up would empower me to work more productively and perhaps open up new career choices. At the very least, I recognized that the four years spent in the programme would break up my routine and challenge me.

Progressing with my doctoral studies, and loving it, I recently realized that I did not see myself becoming a full-time information specialist. So, what could I do? What occupation combines a love for research, a need for sound administration

and people skills, a passion for the Internet and information tools and ties in with politics and international relations? Science policy research — eureka! And so, in early 2004, I applied to the Christine Mirzayan Science & Technology Policy Internship Program of the National Academies, in Washington DC. Lo and behold, this 41-year-old research administrator from Israel was accepted as a summer intern.

My university gave me the time off and I moved to Foggy Bottom, centre of Washington's policy-making. The academies' internship programme taught me many valuable lessons: the city is ripe with opportunity and full of people eager to share their accumulated knowledge and experiences. What struck me most about this town was the enthusiasm and passion that people here attach to their jobs. It is catching. The experience

confirmed for me that science and technology policy research was how I saw myself spending the rest of my professional life.

But, now that I knew what I wanted to do, where could I do it? Israel, my home for 16 years, presents many professional challenges. The university where I work is mid-sized, with an average national science capacity, and strategic planning is not a strong trait. The research office, my professional home for 11 years, is a small department with no mobility opportunities and no mandate to deal with policy issues.

But should I uproot myself, my wife and five children? The experience has taught me that although changing careers is possible, it entails both sacrifices and challenges.

Eric Zimmerman is a grants coordinator at Bar-Ilan University in Israel.

MOVERS

Silke Bühler-Paschen, professor, Technical University of Vienna



Before Silke Bühler-Paschen shone in material sciences, the German physicist was a top-class gymnast. She practised for five hours a day — and for a bit of variety, regularly participated in triathlon competitions.

Perhaps this competitive edge prepared her for the male-dominated physics arena. "I never felt disadvantaged or discriminated against as a woman throughout my career," she says.

Next May she will become the first female full professor of physics at the Technical University of Vienna in Austria — in the city of Ludwig Boltzmann and Erwin Schrödinger.

The charismatic Frenchwoman Marie-Noëlle Bussac, whom she met during her postgraduate training in solid-state physics at the Swiss Federal Institute of Technology in Lausanne, taught her that with discipline, hardiness and dedication one can combine a successful scientific career and a family. Bussac, a theoretical physicist at the École Polytechnique in Paris, and mother of two, became a friend and role model to the young student. Bühler-Paschen now has daughters aged six and three, and is expecting her third child just before she leaves for Vienna next spring.

After a postdoctoral stay in Zurich she won an independent research position at the German Max Planck Society, which has a special programme

to support excellent female researchers. In Dresden, her research has focused on clathrates, cage-like crystals that can trap molecules. Researchers hope to use these compounds to develop special materials with low thermoelectric conductivity, which could replace common cooling agents such as liquid nitrogen. She finds both sides, the basic physical properties of new materials and their potential for application, equally motivating stimuli for her work.

Bühler-Paschen is used to moving around: her father's work in industrial technology took the family to numerous places, including Brazil. Now, a full-time nanny assists her young family at home and during research stays such as one in Japan. And although she no longer has time for gymnastics championships, running and dancing provide a link with her sporting past — and perhaps keep her honed to compete with the rest of the physics community.

CV 2003– Assistant professor, Max Planck Institute for Chemical Physics of Solids, Dresden

1999–2003: Group leader, Max Planck Institute for Chemical Physics of Solids, Dresden

1995–1998: Postdoctoral researcher, Solid State Physics Laboratory, Swiss Federal Institute of Technology, Zurich

1992–1995: PhD researcher, Semicrystalline Solids Physics Laboratory, Swiss Federal Institute of Technology, Lausanne